

How the AI calculates the CEFR level

‘I’m a native speaker of English, so why is my grade less than C2?’

The Dynamic Speaking Test uses AI to evaluate candidates based on their input against five key criteria: Pronunciation, Fluency, Grammar, Vocabulary and Task Achievement. It is crucial to understand that the AI measures only the input provided, and evaluates it without bias.

The standardised evaluation is divided into two main categories: content-based and language-based appropriateness. The assessment criteria include:

- **Pronunciation:** Clarity of speech and proper intonation; making speech intelligible.
- **Fluency:** Maintenance of a natural speech flow without excessive pauses.
- **Grammar:** Production of grammatically correct sentences.
- **Vocabulary:** Use of an appropriate and broad range of vocabulary.
- **Task Management:** Ability to handle various types of communicative tasks appropriately.

Pronunciation: The AI evaluates pronunciation by analysing pronunciation features, flexible use of features and level of intelligibility: Did the AI understand everything that was said? The test should be taken in a quiet environment with a clear microphone to ensure the AI understands all spoken input.

Fluency: This measures the speaker’s ability to maintain a flow of natural speech without excessive pauses. Research shows a strong correlation between fluency, the frequency of pauses and language proficiency.

Grammar: The AI looks for structural accuracy appropriate to each CEFR level. Higher scores require the demonstration of greater grammatical accuracy, using a wide range of complex grammatical structures; errors should be rare and minor.

Vocabulary: Test takers should aim to demonstrate a good range of English vocabulary. Frameworks such as the Cambridge English Vocabulary Profile or the Oxford 3000 and 5000 wordlists can help test takers and educators understand the requirements at different CEFR levels. For advanced levels, a good command of both academic and general English vocabulary is expected, and use of idiomatic expressions and colloquialisms is also necessary.

Task Management: Tasks in the Dynamic Speaking Test require responses that are relevant and adequately address the question. Off-topic responses or those that fail to meet the task requirements result in lower scores.

In summary, the score is determined by the test taker’s input and the AI’s evaluation of it against the assessment criteria. The AI does not make subjective assumptions or infer anything beyond what is explicitly spoken.

Furthermore, unlike human raters, AI does not bring cultural bias into the scoring process, which can lead to inconsistent results influenced by factors like the test taker’s accent, gender, or other personal characteristics. The Dynamic Speaking Test is designed to be free from such biases, promoting fairness and consistency in scoring.

