

How TOEFL Speaking Is Scored: What the Data Shows

Runtian He, CEO, LingoLeap · Published 14 September 2026 · Version 1.0

THREE FINDINGS

- 1 All three dimensions are higher in higher bands. From the 2.0 to the 4.5 AI band, elaboration rises about 1.8 to 4.5 and Language Use 2 to 4.4 (0-5), and fluency 50 to 81 (0-100).
- 2 Topic Development is measurable. Topic relevance runs 2.2 to 5 across the same bands: high answers do not just sound better, they say more.
- 3 Among the delivery signals, fluency has the widest spread (50 to 81), against prosody 67 to 81. The dimensions rise together, so the data cannot say which one drives the score.

CLASSROOM DRILL

Three listeners, three dimensions

15 minutes, groups of four

1. One student answers an interview-style question for about 45 seconds.
2. Each listener judges one dimension only: Delivery (does it flow and sound clear?), Language Use (words and grammar), Topic Development (does it answer the question and explain it with a reason or example?).
3. Each listener gives one strength and one fix for their dimension; rotate until everyone has spoken.

READ THIS BEFORE YOU USE IT

54,247 Take an Interview answers in the 2.0-4.5 bands, from 14,884 LingoLeap practice sessions since January 2026. Delivery signals come from automatic speech assessment and scores from LingoLeap's AI grader; neither is ETS scoring, and both describe LingoLeap practice data only. The dimension names borrow TOEFL's public framework; the numbers are observational associations, not scoring rules.

Source: lingoleap.ai/research/toefl-speaking-scoring · Data: CC BY 4.0 (aggregate numbers only)

Cite: He, R. (2026, September 14). How TOEFL Speaking is scored: What the data shows (Version 1.0). LingoLeap.
<https://lingoleap.ai/research/toefl-speaking-scoring>