

TEST FACT SHEET

English (upper intermediate/B2) Test

The English (upper intermediate/B2) test evaluates a person's knowledge of the English language at the B2 level of the CEFR framework. This test will help you identify people who can participate in upper intermediate professional and social conversations in English.



Covered skills

- ✓ Grammar & Vocabulary
- ✓ Sentence Composition
- ✓ Reading Comprehension
- ✓ Listening Comprehension



Test type

Language



Available languages

English, Dutch, French, German, Spanish, Portuguese (Brazil), Japanese, Italian, Swedish, Danish, Norwegian, Polish



Administration time

10 minutes



Level

Intermediate



Number of questions

20 questions delivered to test-takers
96 questions in the question bank



Use the English (upper intermediate/B2) test to hire

Any person that needs to be highly proficient in English to perform well in their role.



Scoring benchmarks

Benchmarks are available for various education levels (ranging from some high school education to Master's degree or higher), business functions (from administrative to software development), and seniority levels (junior to senior).



Psychometric properties

Each of the metrics reported below is based on a sample size (N) of at least 1,000 candidates, unless indicated otherwise.

Reliability Cronbach's alpha coefficient = .81

Face validity Candidates rated this test as accurately measuring their skills (average score of 4.36 out of 5.00).

Criterion-related validity Candidates with higher scores on this test received higher average ratings from the hiring team during the selection process ($r = .15, N = 668$).

	SUFFICIENT DATA AVAILABLE TO CONDUCT ANALYSES AND CHECKS	ANALYSES AND CHECKS CONDUCTED	ACCEPTABLE OUTCOME
RELIABILITY AND VALIDITY			
RELIABILITY	●	●	●
CONTENT VALIDITY	●	●	●
FACE VALIDITY	●	●	●
CONSTRUCT VALIDITY	●	●	●
CRITERION-RELATED VALIDITY	●	●	●
GROUP DIFFERENCES			
AGE DIFFERENCES	○	○	○
GENDER DIFFERENCES	●	●	●
ETHNICITY DIFFERENCES	○	○	○

○ Pending ◐ Minimally acceptable ● Yes/Good



Glossary

Reliability	The extent to which test scores are stable, consistent, and free from measurement error. Reliability coefficients between .6 and .69 are typically considered reasonable, values between .7 and .79 are considered acceptable, values between .8 and .89 are considered good, and values above .9 are considered great.
Validity	The accuracy of the inferences or interpretations drawn from test scores. There are several types of validity detailed below.
Face validity	The extent to which a test appears to measure what it is intended to measure, and whether, on the surface, the test feels relevant and appropriate for what it is supposed to be assessing. After completing a test, TestGorilla surveys candidates about the perceived validity and relevance of the test.
Content validity	The extent to which a test covers a representative sample of the skills and knowledge content relevant to the topic in question. TestGorilla uses a standardized test development process and formal test structures to ensure the skills and knowledge necessary for a particular topic are well-represented by the test and the test items.
Construct validity	The extent to which the test accurately measures the construct it is intended to measure.
Convergent validity	A type of construct validity. Convergent validity examines whether constructs that are supposed to be theoretically related to each other are, in fact, related. This is the opposite of discriminant validity.
Discriminant validity	A type of construct validity. Discriminant validity examines whether tests that are not supposed to be theoretically related are, in fact, unrelated. This is the opposite of convergent validity.
Criterion validity	The degree to which test scores are related to scores on an outcome measure of interest (e.g. performance ratings, turnover). Based on the newest research in the field, values of validity coefficients below .07 are typically seen as low, values between .07 and .15 are seen as moderate, values between .16 and .29 are seen as substantial, and values above .29 are considered high.
Group differences	The extent to which different groups (e.g. different age, gender, ethnic and/or racial groups) differ significantly from each other in terms of the scores obtained on a test.