# TESTING & EXAMS IN ELT

What makes a good test, how to assess speaking and writing, and how to use assessment to support learning

AT A GLANCE

| | |
|---|---|
| **What it is** | finding out what learners know and can do in English, through tests, exams and everyday classroom assessment |
| **Key ideas** | assessment **of** learning (measuring it) and assessment **for** learning (supporting it); validity, reliability and washback |
| **Key message** | good assessment measures what matters, is fair, and has a positive effect on teaching and learning |
| **Tools** | a **test design checklist** and a **sample speaking rating scale** in the appendices |

THINK ABOUT

- Think of a test you took as a language learner. Did it show what you could really do in the language?
- How much does an exam your learners take influence what and how you teach?
- Apart from tests, how do you find out whether your learners are making progress?

## Assessment, testing and evaluation

These terms are often used interchangeably, but they mean different things.

| Term | Meaning |
|---|---|
| **Assessment** | any way of gathering information about what learners know and can do: tests, observation, homework, portfolios, self-assessment |
| **Testing** | one form of assessment: a planned procedure, usually at a particular time, with tasks and scores |
| **Evaluation** | judging the effectiveness of a course, programme or materials, often using assessment results |

### Summative and formative assessment

| | Summative assessment | Formative assessment |
|---|---|---|
| **Purpose** | to measure what learners have achieved, often for a grade or certificate | to find out where learners are, so that teaching and learning can be adjusted |
| **When** | usually at the end of a unit, course or level | throughout the course, often during lessons |

| | Summative assessment | Formative assessment |
|---|---|---|
| **Examples** | end-of-term tests, international exams | questions, observation, mini-quizzes, self-assessment, feedback on drafts |
| **Also called** | assessment **of** learning | assessment **for** learning |

The same task can be summative or formative: what matters is how the results are used.

## Types of tests

Hughes (2003) groups tests according to their purpose.

| Type | Purpose | Example |
|---|---|---|
| **Placement** | to put learners in the right class or level | a test and short interview when learners join a school |
| **Diagnostic** | to identify learners' strengths and weaknesses | a test at the start of a course to plan priorities |
| **Progress** | to check learning during a course | a short test at the end of each unit |
| **Achievement** | to measure what learners have learned by the end of a course | an end-of-course test based on the syllabus |
| **Proficiency** | to measure general ability in the language, regardless of any course | international exams such as IELTS, TOEFL or Cambridge English Qualifications |

### Other useful distinctions

- **Direct and indirect tests:** a direct test asks learners to perform the skill (e.g. write a letter); an indirect test measures something related to it (e.g. choosing correct sentences).
- **Discrete-point and integrative tests:** discrete-point items test one element at a time (e.g. a single verb form); integrative tasks combine several (e.g. writing a summary of a text).
- **Norm-referenced and criterion-referenced tests:** norm-referenced tests compare learners with each other; criterion-referenced tests compare them with a standard, e.g. CEFR descriptors.

## What makes a good test?

Brown & Abeywickrama (2019) describe five key principles.

| Principle | What it means | Ask yourself… |
|---|---|---|
| **Validity** | the test measures what it is meant to measure | Does a speaking test really require speaking? Does a progress test cover what was taught? |
| **Reliability** | results are consistent: the same learner would get a similar score on another day or with another marker | Are instructions clear? Are marking criteria clear and used consistently? |
| **Practicality** | the test is realistic in terms of time, cost and resources | Can I give and mark this test with the time and people I have? |

| Principle | What it means | Ask yourself… |
|---|---|---|
| **Authenticity** | tasks resemble real-world language use | Would learners ever need to do something like this outside the test? |
| **Washback** | the effect of the test on teaching and learning | Will preparing for this test help learners learn useful English? |

*Based on Brown & Abeywickrama (2019)*

These principles sometimes pull in different directions. A face-to-face speaking test is valid and authentic, but less practical for a class of 50 than a written test. Good test design involves balancing them.

### Washback

Tests influence what happens in the classroom. **Negative washback** occurs when teaching narrows to what the test covers: for example, practising multiple-choice items instead of developing speaking. **Positive washback** occurs when preparing for the test also develops useful skills (Alderson & Wall 1993). Hughes (2003) suggests encouraging positive washback by testing the abilities you want to develop, using direct tests, and making sure learners and teachers understand what the test requires.

## The CEFR

The **Common European Framework of Reference for Languages** (CEFR) describes language ability in six levels, from A1 to C2, using "can-do" descriptors. It is now used around the world to describe courses, materials and exams. The Companion Volume (Council of Europe 2020) updated the descriptors and added areas such as mediation.

| Level | In brief, learners can… |
|---|---|
| **A1** | understand and use familiar everyday expressions and very basic phrases |
| **A2** | communicate in simple, routine tasks on familiar topics |
| **B1** | deal with most situations when travelling, and describe experiences and plans |
| **B2** | interact with some fluency and spontaneity, and understand the main ideas of complex texts |
| **C1** | use language flexibly and effectively for social, academic and professional purposes |
| **C2** | understand virtually everything heard or read, and express themselves precisely |

*Summarised from the CEFR global scale (Council of Europe 2020)*

The CEFR is a useful common reference, but it is not a test. Teachers can use its descriptors to set goals, design tasks, and help learners assess themselves.

## Assessing speaking and writing

Speaking and writing are usually assessed with **rating scales**, which describe performance at different levels.

| | Holistic scale | Analytic scale |
|---|---|---|
| **What it is** | one overall score for the whole performance | separate scores for different criteria |

| | Holistic scale | Analytic scale |
|---|---|---|
| **Advantages** | quick to use; reflects the overall impression | more detailed; more reliable; gives useful feedback |
| **Disadvantages** | less detailed; harder to explain a score | takes longer to use |

Typical criteria include **task achievement** or **content**, **organisation** or **discourse management**, **range and accuracy** of grammar and vocabulary, **pronunciation** (for speaking) and **interaction**. Appendix B gives a sample analytic scale for speaking.

### Making marking more reliable

- use clear criteria and share them with learners before the task
- **standardise**: agree on scores for sample performances with colleagues before marking
- **double mark** a sample of performances and discuss differences
- record speaking tests where possible, so you can check your marks later

## Writing good test items

| Item type | Good for | Watch out for |
|---|---|---|
| **Multiple choice** | reading and listening comprehension; grammar and vocabulary; quick, reliable marking | hard to write well; learners can guess; tests recognition, not production |
| **Gap-fill** | grammar and vocabulary in context | make sure only one answer fits, or accept all correct answers |
| **Matching** | vocabulary; matching headings to paragraphs | include extra options so the last answer isn't automatic |
| **True / false** | quick checks of understanding | a 50% chance of guessing; add "not given" or ask learners to correct false statements |
| **Short answer** | comprehension; focused production | decide in advance which answers are acceptable |
| **Extended writing** | writing ability | needs clear criteria and careful marking |
| **Interview or role-play** | speaking ability | time-consuming; needs clear criteria and, ideally, two assessors |

**TEACHING TIP**

When writing multiple-choice items, make sure there is only one correct answer, that all the wrong options (distractors) are plausible, and that the correct answer isn't given away by its length or by grammar clues.

## Assessment for learning in the classroom

Research shows that formative assessment, used well, can significantly improve learning (Black & Wiliam 1998). Wiliam (2011) describes five key strategies.

| Strategy | In the ELT classroom |
|---|---|
| **Clarify learning intentions and success criteria** | tell learners what they are learning and what a good performance looks like, e.g. with a model or checklist |
| **Find out what learners know** | use questions, mini-whiteboards, polls and short tasks to see every learner's understanding |
| **Give feedback that moves learning forward** | tell learners what they did well and what to do next, rather than just giving a mark |
| **Use learners as resources for each other** | peer feedback with clear criteria; learners checking and explaining to each other |
| **Help learners own their learning** | self-assessment with can-do statements; learning journals; setting personal goals |

*Based on Wiliam (2011)*

## Teaching exam classes

Many learners study English for an international exam, such as Cambridge English Qualifications, IELTS, TOEFL iBT, PTE or Trinity exams. Exam formats change, so always check the latest official information from the exam board.

| Do… | Avoid… |
|---|---|
| make sure learners know the format, timing and assessment criteria | only doing practice tests, lesson after lesson |
| develop the skills and language the exam tests | teaching tricks instead of language |
| teach useful strategies, e.g. predicting, managing time, checking work | ignoring skills that the exam doesn't test, if learners need them |
| use official sample materials and marking criteria | relying on unofficial materials that may not match the exam |
| include timed practice gradually, and give feedback on it | adding unnecessary pressure: help learners manage anxiety |

## Assessment and AI

AI tools can help teachers write test items, create rating scales and draft feedback, but every item and comment needs checking. They also make it easier for learners to submit work they didn't write, and AI detection tools are unreliable. Assessment that values the **process**, such as drafts, in-class writing, oral follow-up questions and reflection, is more robust and more useful for learning (see the **Technology in ELT** handout).

# Making it work

## Start from what you want learners to do

Decide what learners should be able to do, then choose a task that shows it.

## Share criteria with learners

Learners do better when they know what a good performance looks like.

## Use results

Use what you learn from assessment to plan teaching and give feedback, not only to give grades.

# Common pitfalls

| Pitfall | What to do instead |
|---|---|
| **Testing only grammar and vocabulary.** | Include speaking, writing, reading and listening. |
| **Testing something different from what was taught.** | Base progress and achievement tests on the course. |
| **Marking speaking and writing on impression alone.** | Use clear criteria and standardise with colleagues. |
| **Giving marks without feedback.** | Tell learners what they did well and what to do next. |
| **Turning exam classes into endless practice tests.** | Develop skills and language, with practice tests as one part. |

ONLINE CLASSES

Online quizzes and forms make progress tests quick to give and mark. For speaking, record learners on video calls or ask them to submit recordings. Remember that online tests are harder to supervise: design tasks that show learners' own thinking, such as follow-up questions.

KEY TAKEAWAYS

- Assessment can measure learning (summative) or support it (formative).
- Good tests are valid, reliable, practical and authentic, and have positive washback.
- The CEFR provides a common reference for levels and can-do goals.
- Assess speaking and writing with clear criteria, and standardise marking.
- In exam classes, develop real skills, not just exam techniques.

WHERE NEXT

**Error Correction** feedback during and after tasks

**Teaching Speaking and Teaching Writing** what speaking and writing involve

**Mixed-Ability Classes** recognising effort and progress

**Technology in ELT** online quizzes and AI

## Further reading

Brown, H.D. & Abeywickrama, P. (2019). *Language Assessment: Principles and Classroom Practices* (3rd ed.). New York: Pearson Education.

Green, A. (2014). *Exploring Language Assessment and Testing*. Abingdon: Routledge.

Hughes, A. (2003). *Testing for Language Teachers* (2nd ed.). Cambridge: Cambridge University Press.

Wiliam, D. (2011). *Embedded Formative Assessment*. Bloomington, IN: Solution Tree Press.

## References

Alderson, J.C. & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14(2), 115–129.

Black, P. & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5(1), 7–74.

Council of Europe (2020). *Common European Framework of Reference for Languages: Companion Volume*. Strasbourg: Council of Europe.

# Appendix A · Test design checklist

Use this checklist when you write a progress or achievement test.

| Purpose and content | |
|---|---|
| I know what the test is for: placement, diagnosis, progress or achievement. | ☐ |
| The test covers what was taught, in a balanced way. | ☐ |
| It includes skills as well as grammar and vocabulary. | ☐ |

| Tasks and items | |
|---|---|
| Each task tests what it is meant to test. | ☐ |
| Instructions are short, clear and at a lower level than the test content. | ☐ |
| Each item has one correct answer, or I have listed all acceptable answers. | ☐ |
| Speaking and writing tasks have a clear purpose and audience. | ☐ |

| Marking | |
|---|---|
| I have an answer key and clear criteria for speaking and writing. | ☐ |
| Colleagues and I have agreed how to apply the criteria. | ☐ |
| Learners know the criteria before the test. | ☐ |

| Practicality and impact | |
|---|---|
| The test can be given and marked in the time available. | ☐ |
| Preparing for the test will help learners learn useful English. | ☐ |
| Results will be used to give feedback and plan teaching. | ☐ |

## Appendix B · Sample speaking rating scale

A simple analytic scale for a short speaking task, e.g. a role-play or discussion at B1. Adapt the descriptors to your learners' level and task. Score each criterion from 1 to 3.

| Criterion | 3 | 2 | 1 |
|---|---|---|---|
| **Task achievement** | completes the task fully; ideas are relevant and developed | completes most of the task; some ideas are limited | completes little of the task; ideas are very limited |
| **Range and accuracy** | uses a good range of grammar and vocabulary for the task; errors rarely cause problems | uses adequate language; errors sometimes cause problems | uses very limited language; errors often cause problems |
| **Fluency and organisation** | speaks with little hesitation; ideas are clearly linked | some hesitation; ideas are mostly linked | frequent hesitation; ideas are hard to follow |
| **Pronunciation** | is easy to understand throughout | is generally understood, with some effort | is often difficult to understand |
| **Interaction** | responds naturally, asks questions and keeps the conversation going | responds appropriately but needs some support | responds with difficulty; needs a lot of support |

| Learner | Task | TA | R&A | F&O | P | I | Total / 15 |
|---|---|---|---|---|---|---|---|
| | | | | | | | |
| | | | | | | | |

**Feedback: what went well, and what to work on next**