



We believe fair tests should be as *objective* as possible

Assessment is not the same as judgment. We believe PTE Academic is the most objective test of academic English in the world.

The best tests are objective and therefore give results that can be relied on. It should not matter to the test taker when they take the test, where they take the test, or who scores the test; they should be confident of getting a result which is consistent with their ability.

Likewise score users need to have confidence that a test score accurately reflects the ability of the test taker and can be taken as an indication of how the test taker will cope in an environment where the abilities being tested are called upon.

What this means for you and your institution:

- PTE Academic is machine scored. The responses are therefore scored objectively and consistently, no matter where in the world the test is taken. There are no regional variances in standards meaning you can have complete confidence in the communication skills of your students
- PTE Academic assesses English language skills in more ways than any other test. This means that your students will be more likely to cope well with the wide variety of tasks expected of them at your institution
- It doesn't matter where your students come from or what they look like, PTE Academic only tests their English skills. You can therefore be sure that all PTE Academic test takers have been treated with complete fairness, and that you are not missing out on students because of unreliable test scores
- PTE Academic treats regional accents equally. Human assessors tend to value accents they are used to higher than accents they are unfamiliar with ¹. You can therefore be confident in the PTE Academic scores of international students from all over the world

We demonstrate the objectivity of PTE Academic in the following ways:

1

Reliability estimates

Each of the major tests of academic English reports a reliability estimate. A reliability estimate is expressed as a number between 0 and 1, where 0 means totally unreliable and 1 means perfectly reliable. For tests that are used to make important decisions, high reliability (0.90 or higher) is required. High reliability means that score users can trust the results, i.e., that the results on repeating the test will be the same or very close. How close you can expect repeat results to be is indicated by the reliability estimate: the higher the reliability estimate, the closer the repeat results.

The table below captures the reported reliability estimates for PTE Academic, IELTS and TOEFL iBT. PTE Academic has the highest reliability estimates for both the overall score and the communicative skills scores of all the major academic English tests.

	PTE Academic	IELTS	TOEFL iBT
Overall Reliability	0.97 ²	0.96 ³	0.94 ⁴
Communicative skills reliability			
Reading	0.92	0.90	0.85
Listening	0.91	0.91	0.85
Writing	0.91	0.81-0.89	0.74
Speaking	0.91	0.83-0.86	0.88

This demonstrates that PTE Academic produces consistent, trustworthy results.

Objectivity through automated scoring

An aspect of test fairness is how a test is scored. The major academic English tests use different approaches for assessing test taker performance. This comprises human raters only, human raters combined with machine scoring or machine scoring only. PTE Academic uses machine scoring only.

	PTE Academic ⁵	IELTS ⁶	TOEFL iBT ⁷
Scoring Method: Writing	Automated scoring	Human rating	Human rating Automated rating used for two tasks
Scoring Method: Speaking	Automated scoring	Human rating	Human rating

It is widely recognized that the scores human raters give can be influenced by irrelevant factors, particularly when only one person rates the test taker's performance. Automated scoring has the benefit of removing this effect as it is indifferent to a test taker's appearance and personality, and is not affected by issues such as examiner tiredness, mood and leniency. Automated scoring also allows individual features of a language sample (e.g. vocabulary or pronunciation) to be analyzed independently, so that weakness in one area of language does not affect scoring in others.

Although PTE Academic is computer-based and machine-scored, responses to all new tasks that require constructed responses (e.g. essays) are initially human scored so that the intelligent scoring systems can be appropriately trained and calibrated. PTE Academic only uses well-trained human raters who are retrained and certified before each rating session. And because PTE Academic's automated scoring is based on the collective wisdom of a large pool of skilled human raters (over 200) and not, as in human scoring, on the views of one or a handful of individuals, every single response receives a replicable, objective and completely impartial score.

The most common way of determining the fairness of scoring is to look at how well two or more raters agree on a test taker's score. This comparison may also be established between a human rater and a machine. Research carried out by Shohamy, Gordon and Kraemer⁸ in 1992 showed that agreement between two human raters typically ranged from 80% to 93%, depending on the level of training involved and the scales being used. With PTE Academic, however, the overall agreement between human ratings and machine generated scores is 97%².

In addition, using automated scoring means that the test takers' natural accents are all dealt with objectively. During the development of PTE Academic, close to half a million responses from more than ten thousand test takers originating from 158 countries of birth and speaking 126 different languages at home were used to optimize Pearson's speech recognizer. The speech recognizer was trained to treat all their accents equally using the same criteria. Most human raters are likely to be more accustomed to some accents than others which will make it difficult for them to treat accents they are not familiar with equally. Some may be accustomed to perhaps half a dozen accents but it is extremely unlikely that they will be able to treat over hundred different foreign accents objectively as PTE Academic does.

This demonstrates that PTE Academic results are an objective and unbiased reflection of a test taker's true language skills.

References

- ¹. Carey, M.D., Mannell, R.H., and Dunn, P.K. (2011) "Does a rater's familiarity with a candidate's pronunciation affect the rating in oral proficiency interviews?", *Language Testing*, 28(2), 201-219.
- ². Pearson (2012) Validity and Reliability in PTE Academic
<http://pearsonpte.com/research/Pages/ValidityandReliability.aspx>
- ³. IELTS Partners (2011) IELTS | Researchers - Test performance 2011
www.ielts.org/researchers/analysis_of_test_data/test_performance_2011.aspx
- ⁴. ETS (2011) Reliability and Comparability of TOEFL iBT™ Scores
www.ets.org/s/toefl/pdf/toefl_ibt_research_slv3.pdf
- ⁵. Pearson (2011) PTE Academic and Automated Scoring
<http://www.pearsonpte.com/research/Pages/AutomatedScoring.aspx>
- ⁶. IELTS Partners (n.d.) How is IELTS Marked?
http://www.ielts.org/institutions/test_format_and_results/how_is_ielts_marked.aspx
- ⁷. ETS (2012) TOEFL iBT: Scores
<http://www.ets.org/toefl/ibt/scores/>
- ⁸. Shohamy, E., Gordon, C.M., & Kraemer, R. (1992) The effects of raters' background and training on the reliability of direct writing tests. *The Modern Language Journal*, 76 (1), 27-33

*IELTS is a registered trademark of the University of Cambridge ESOL Examinations Syndicate.
TOEFL iBT is a registered trademark of Educational Testing Service (ETS).*