

Test Validity and Secure Test Delivery

David Booth



Test Construct

- A measure of the English language ability of test takers needed to actively participate in tertiary level education where English is the language of instruction

Secure Test Delivery

- Both test integrity and security are amongst Pearson's highest priorities for PTE Academic. They are integral to all aspects of the program, from test development to delivery. In addition, physical security at testing centers, particularly in terms of test takers' identification, is of paramount importance when developing a high-quality test.

Test Integrity

- **Test integrity** is the quality that gives users of a test confidence that the results they receive are genuine and represent a true depiction of the ability of the test taker in question.

Test Security

- **Test security** is focused on protecting the confidentiality of test content and ensuring that tests are administered in such a way that test takers are not able to gain an unfair advantage through cheating.

UKBA Requirements - Test Delivery

- Evidence that tests are delivered in a secure testing environment
- Evidence of the processes for the control and storage of test registration/test papers/test material to ensure test security and prevent unauthorised access.
- Evidence of the processes for the movement and handling of test papers/test material on test days to ensure test security and prevent unauthorised access.
- Evidence of security arrangements where computer-based assessment is used, in particular the prevention of access to other computer based tools such as the internet.

UKBA Requirements - Test Delivery and Awards

- Evidence that appropriate, standardised invigilation arrangements are in place
- Evidence that appropriate, standardised test marking arrangements are in place
- Evidence that test providers audit their test delivery processes and that these records are filed and available for inspection
- Evidence that their test(s) awards are secure and contain a number of security features to counteract any attempts to produce fraudulent evidence.

Main Security Risks

- ‘Old fashioned’ cheating
- Imposters/Ringers/Proxy testing
- Fraud
 - Certificates/Test Report Forms
 - Test centre issues
 - Systematic and organised cheating

Measures to Combat Risks

- Test development
- Test publication
- Data transfer
- Test taker and test data protection
- Test centre security
- Scoring and reporting results

Test Centre Security

- Test premises
- Test taker identification checks
- Biometrics
- Delivery workstation security

Test Centre Security

- Pearson uses a variety of biometrics including digital photographs, fingerprints, palm-vein scans, and signatures to verify the identity of the test taker. Test center Administrators also utilize constant audio and video monitoring and recording in the test center to provide the highest level of security.

Test Centre Security - Proxy Testing

- In addition to the identification requirements, each test taker is required to provide three digital biometrics: a signature, photograph, and fingerprint or palm-vein scan. The signature, photograph, and other personal data on file are compared to the test taker's ID to ensure an exact match. This information is also compared against other test takers' data to ensure that the same individual is not testing under multiple registrations.

Scoring and Reporting Results

- Machine scoring
- Result holds
- Data forensics
- Results service

Scoring and Reporting Results

- In many ways, automated scoring gives more analytical and objective results than human scoring does. Unlike human judgment, which is prone to be influenced by extraneous factors, an automated scoring system is consistent and impartial.

Scoring and Reporting Results

- Automated scoring also allows individual features of a language sample (spoken or written) to be analysed independently, so that weakness in one area of language does not affect the scoring of other areas.

What does it mean to be “at a level”?

Being at a CEF level means:

Being more likely than not be able to perform language activities at the level of efficiency such as captured in the exemplary descriptors provided in the CEF.

It CANNOT be

Being able to perform ALL language activities for a particular level provided in the CEF.

Simply because the CEF only contains exemplary descriptors.

Test Development - Mapping to the CEF

Four methods

- Item writer training & writing to CEF
- Item estimates of CEF level in reviewing process
- Inclusion of DIALANG anchors in field test
- Human rating on CEF scale of candidates

Two independent estimates of CEF cut-offs

