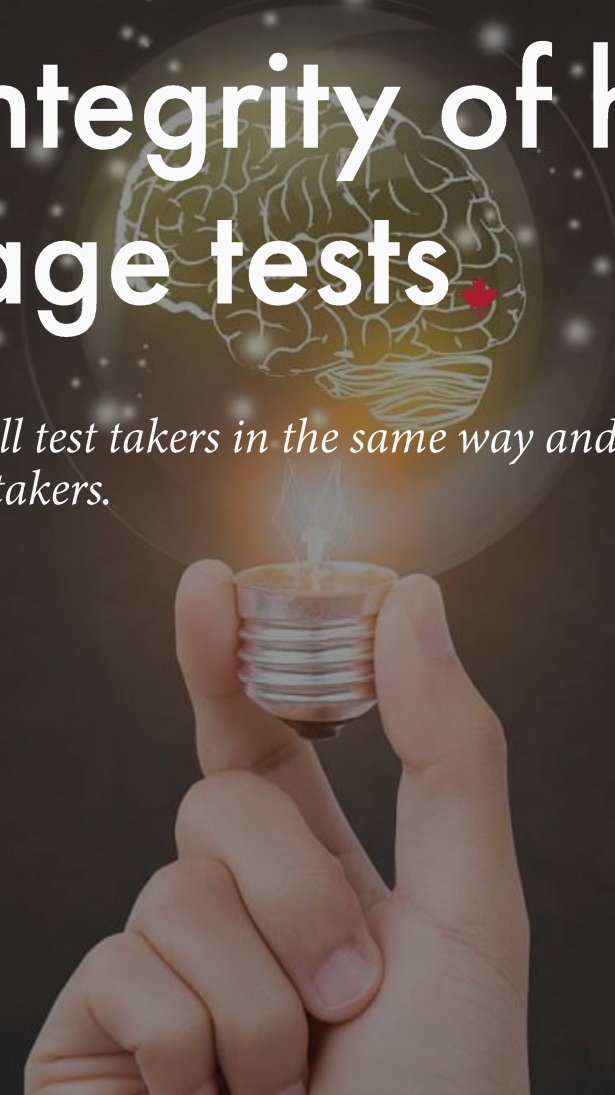


Testing for fairness: Ensuring the quality and integrity of high-stakes language tests



There is a difference between assessing all test takers in the same way and making sure assessments are appropriate and inclusive of all test takers.

Fairness

Fair = free from bias, dishonesty, or injustice

Synonyms = impartial, unprejudiced

“Fair” implies treating all sides alike, justly, and equitably

What may make a test fair or unfair?

- Scores for some people or groups are lower than expected
- The test is too hard
- The test is a barrier to prevent something from happening
- Content was biased
- Test takers did not have an equitable testing experience
- Equal treatment of test takers
- Complexity of questions appropriate to the purpose of the test
- Clear instructions
- Good test security
- Accessible testing facilities
- Lack of bias

Stakeholders in high-stakes testing

Test professionals: The individuals who design and administer the test

Test score users: The organizations or individuals that use the information from the test to make decisions

The test takers: The individuals who take the test or are preparing to take the test

Fair Testing

- Fairness is complex
- Fairness is a validity issue
- Fairness is the absence of bias

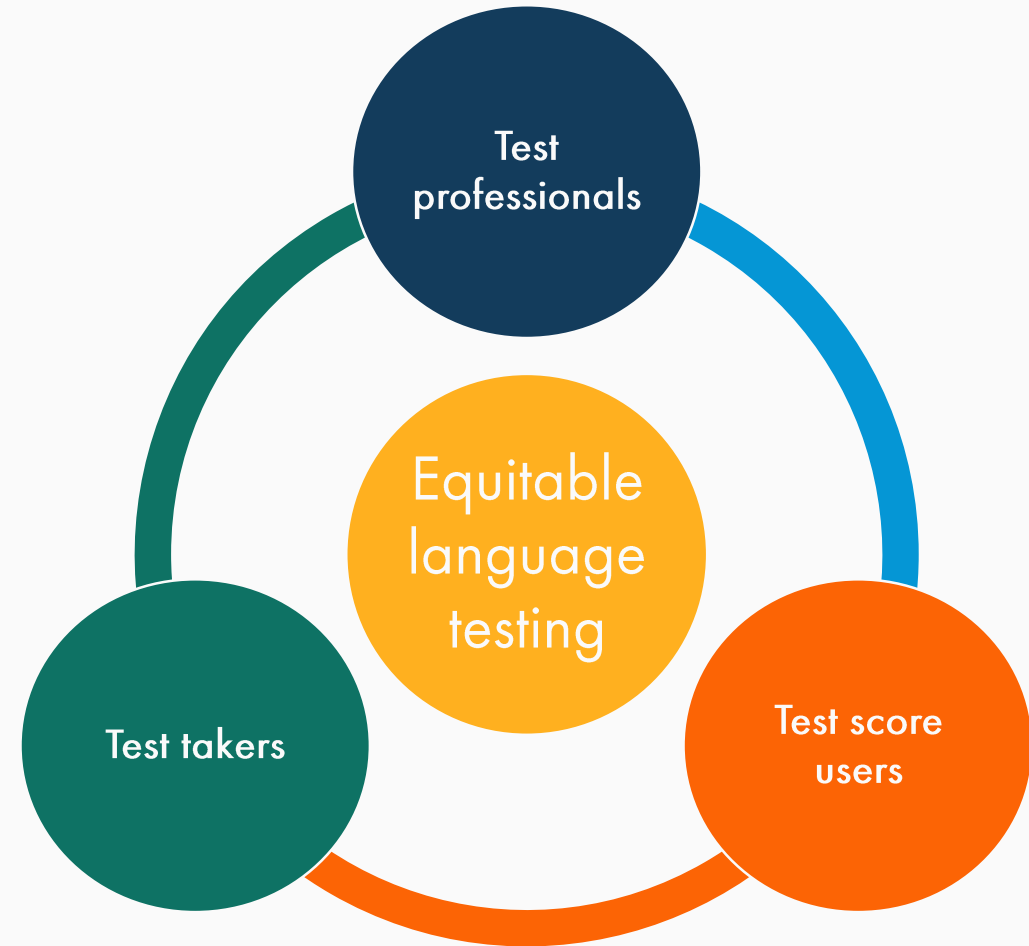
The goal:

- Tests reflect the same constructs for all test takers
- Test conditions are replicated for all test takers as they are set out in the delivery specifications
- Scores have the same meaning for all individuals in the intended population

Equitability in Testing

Equitability implies joint responsibility.

Test takers must be given a comparable opportunity to demonstrate their skills and abilities



Role of test professionals

- Educate and inform test users and test takers
- Justify any claims made about the test
- Encourage participation from test users and test takers
- Solicit feedback from test users and test takers

Test professionals

1. Test design
2. Test development
3. Test administration
4. Test scoring
5. Test research

“Fairness is a fundamental validity issues and requires attention throughout all stages of test development and use” (*Standards*, p. 49)

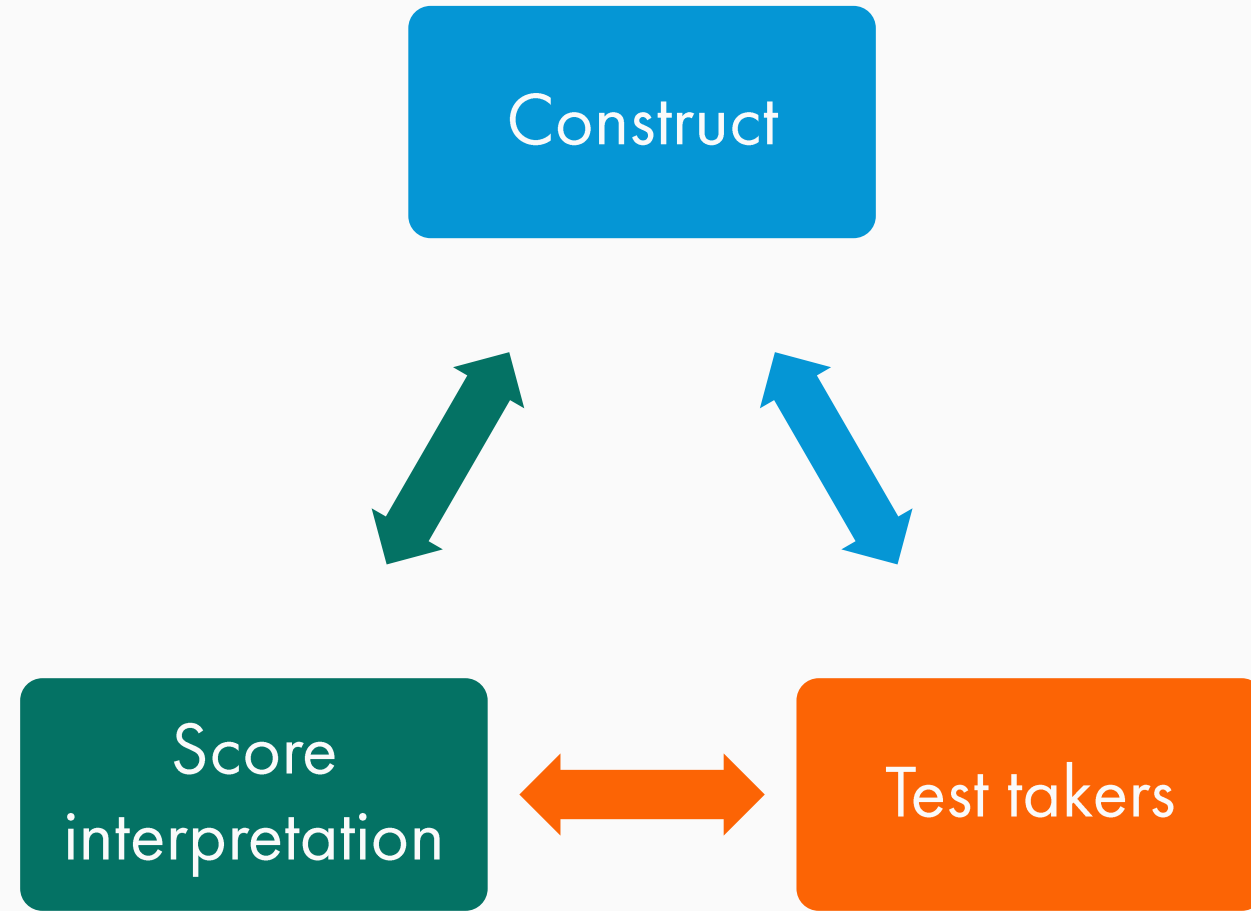
1. Test design

Testing only construct-relevant knowledge (i.e., language knowledge).

The knowledge and skills that a test **intends** to measure is referred to as its construct.

Construct relevant knowledge and skills are required for reading, listening, speaking and writing in English.

Construct-irrelevant knowledge and skills are not related to language proficiency.



2. Test development

Test development process

- Research and investigation (construct, domain, best practice, etc.)
- Recruit and train item writers
- Develop a sufficient number of test items to create a pool from which parallel forms can be assembled
- Review the test items, applying standards of fairness and quality
- Try-out (field test) items
- Analyze and review test data to verify the items are functioning as intended in terms of scoring and fairness to all groups

2. Test development (continued)

- Writing content that is sensitive to a diverse test taking population.
- Avoiding content that advantages/disadvantages certain groups of test takers.
- Excluding topics that may create an emotional barrier to a test taker's performance and handling sensitive topics with care.
- Reviewing questions qualitatively and quantitatively for bias.

Creating content that is free from bias provides equal opportunities for all test takers to show their skills and abilities.

3. Test administration

- Training and monitoring of test administrators.
- Adhering to unified test administration procedures during all test administrations.

Following administration procedures ensure that differences in performance are related to language proficiency, rather than to irrelevant factors.

4. Test scoring

A. Role of test professionals

Listening and Reading items

- Score equating: A statistical procedure that ensures score are comparable across test forms and dates

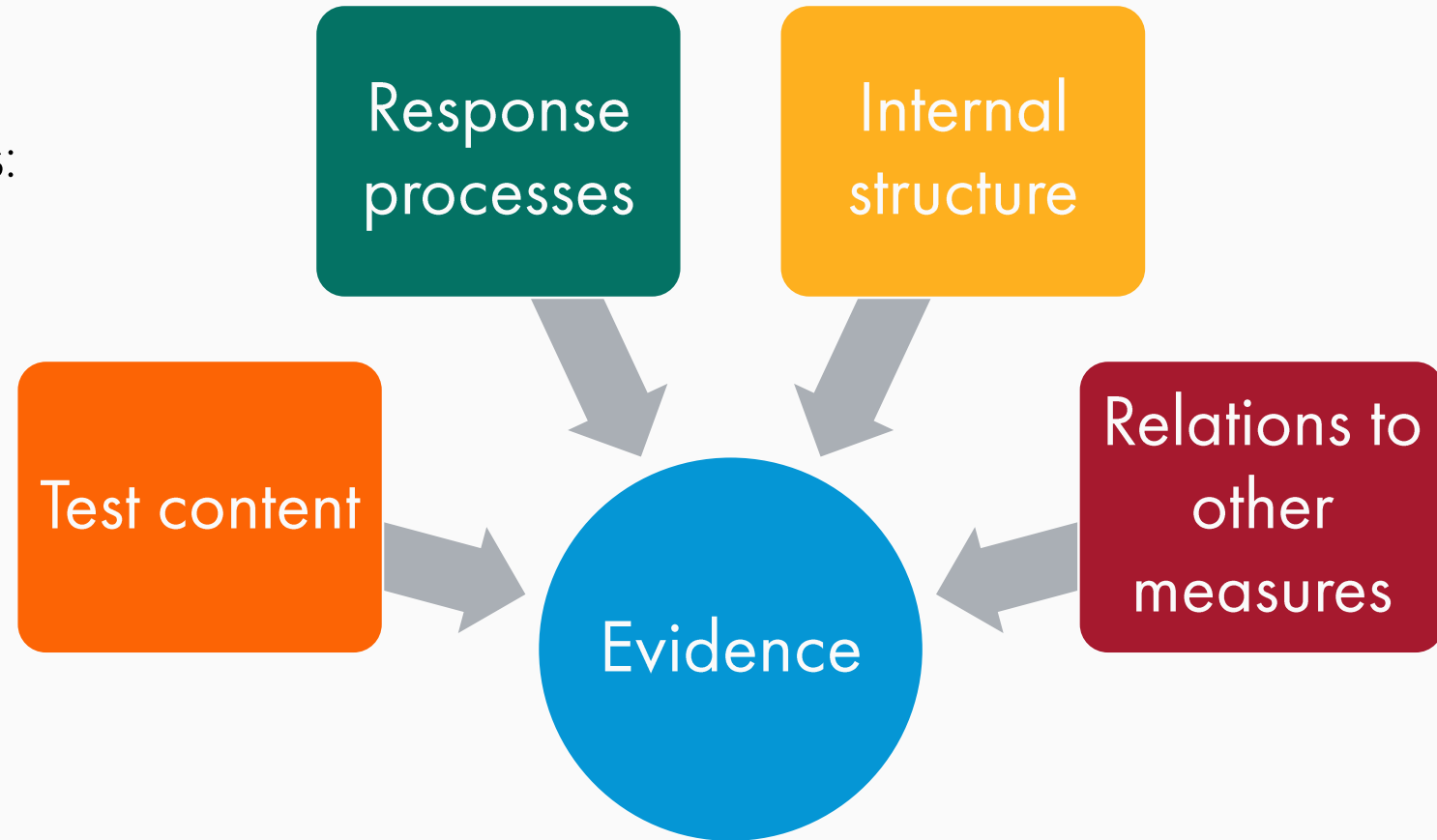
Speaking and Writing tasks

- Selecting raters who have the right qualifications, attitude, outlook, and knowledge to assess test takers.
- Initial and ongoing training and re-certification of raters.
- Continual monitoring of rater quality.

5. Test research

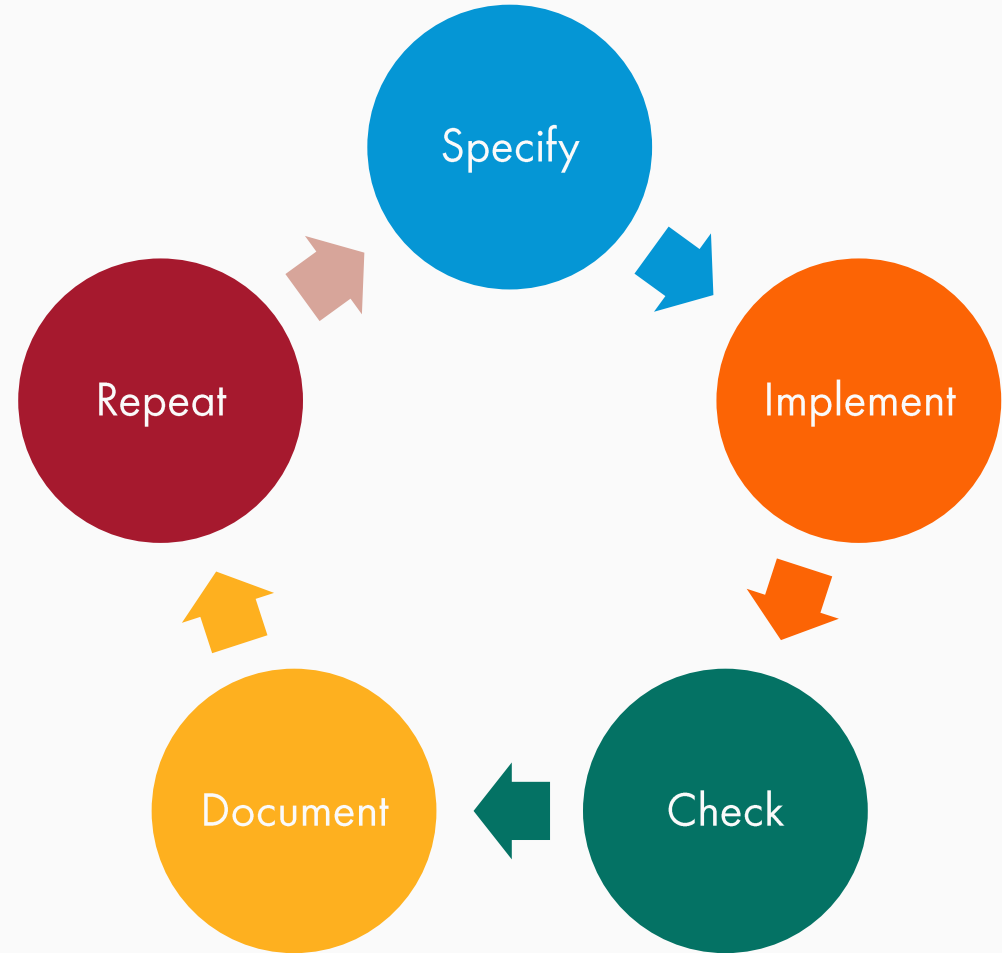
Evidence can take one of two forms:

1. Procedural
2. Empirical



Test professionals' role summarized

Clearly describe the construct being measured, explain the intended use of the test, and create and implement specifications and procedures that allow test users to make valid inferences.



Role of test users

Test users

1. Apply the right test to the right population
2. Use appropriate test scores to make decisions
3. Evaluate procedures used by testing organizations
4. Review the performance of test takers

Test utility

The degree to which a test improves the quality of a selection system, above and beyond what would have occurred has the test not been used.

Role of test takers

Test takers

1. Inform themselves about the test content and test method beforehand
2. Match the level of the test to their skill level

Conclusions

Equitable experience for all

Equitable treatment

- All test takers are given a comparable opportunity to demonstrate their standing on the construct(s)
- Appropriate testing conditions
- Equal opportunity to become familiar with the test format, practice materials, etc.
- Appropriate and fully informative score reporting

Equality of outcomes

- Test takers who have the same level of language proficiency should achieve a similar score.

Thank you

Lauren Kennedy-Barrows

Senior Product Development Specialist | Assistant Manager, Test Research & Development

Email: barrows@paragontesting.ca

Visit: paragontesting.ca