Putting tests to the test

Six questions all teachers should ask

Evelina Galaczi and Nahal Khabbazbashi
IATEFL 2016
Birmingham, UK
Our aim today

To help you to:

• UNDERSTAND assessment

• DEVELOP tests which support learning

• be able to EVALUATE existing tests.
Key concepts in assessment

Validity

Is the test measuring what it claims to be measuring?

- Test PURPOSE
- Test TAKERS
- Test CONSTRUCT
- Test TASKS
- Test RELIABILITY
- Test IMPACT
1. Why am I testing?

TEST PURPOSE

A. To CHECK LEARNING at the end of a unit in school?
B. To DIAGNOSE what learners know and don’t know?
C. To PLACE learners into groups based on their ability?
D. To provide test takers with a CERTIFICATE of language proficiency?

Six key questions
2. Who am I testing?

Six key questions

TEST TAKERS

• Primary school children?
• Teenagers?
• Adults?
• Airline pilots?
• Doctors?
3. What am I testing?

Six key questions

TEST CONSTRUCT

• Communicative language ability?
• Speaking ability?
• Pronunciation?
• Grammatical knowledge?
• Use of grammatical knowledge?
Test construct

The ability a test tries to measure = test CONSTRUCT
(Ability = latent trait)

For example:
- maths
- intelligence
- personality
- anxiety
- English language ability
- pronunciation
Six key questions

4. How am I testing the construct?

TEST TASKS

• Grammar and vocabulary -> multiple choice?
• Reading ability -> texts and questions?
• Listening ability -> a lecture and questions?
• Writing ability -> a short email?
• Speaking ability -> face-to-face interview?
5. How am I scoring the test?

Six key questions

RELIABILITY

• Correct/incorrect?
• Using trained teachers?
• Using assessment criteria?
6. How is my test benefitting learners?

TEST IMPACT

• Through using AUTHENTIC tasks?
• Through including all FOUR SKILLS (reading, writing, listening, speaking)?
• Through formative FEEDBACK on performance?
To summarise

Six key questions
1. Why am I testing?
2. Who am I testing?
3. What am I testing?
4. How am I testing?
5. How am I scoring?
6. How is my test benefiting learners?

Six key assessment concepts
a. Test PURPOSE
b. Test TAKERS
c. Test CONSTRUCT
d. Test TASKS
e. Test RELIABILITY
f. Test IMPACT
Validity

• Is the test measuring what it claims to be measuring?

• A test is never just valid, it is VALID FOR a purpose

→ FITNESS FOR PURPOSE
In the rest of the presentation:

Six key questions
• Why am I testing?
• Who am I testing?
• What am I testing?
• How am I testing?
• How am I scoring?
• How is my test benefiting learners?

Six key assessment concepts
✓ Test PURPOSE
✓ Test TAKERS
✓ Test CONSTRUCT
✓ Test TASKS
✓ Test RELIABILITY
✓ Test IMPACT
Construct = a cognitive element
Construct = a task element

What are the cognitive processes required to complete the task?
Cognitive processes in practice

Read the sentence and answer the question.

Yesterday I saw the palgish flester gollining begrunt the bruck.

**Question 1:** What was the flester doing?

A. chandering
B. gollining
C. rangeling

Cognitive processes in practice

Yesterday I saw the palgish flester gollining begrunt the bruck.

Question 2: Where was the flester?

A. begrunt the quistly
B. besand chander
C. begrunt the bruck

Yesterday I saw the palgish flester gollining begrunt the bruck.

Question 3: What is the event described here?
A. The flester is participating in a sports competition.
B. The flester is cooking dinner for friends.
C. The flester is ill in hospital.

Cognitive processes in practice

What makes Questions 1 and 2 different from Question 3?
What makes Questions 1 and 2 different from Question 3?

- Questions 1 and 2 → sentence-level knowledge
- Question 3 → making inferences
Cognitive processes in practice

<table>
<thead>
<tr>
<th>Careful reading</th>
<th>Local level</th>
<th>Global level</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Understanding LEXICAL and GRAMMATICAL meaning</td>
<td>Understanding MAIN IDEAS and supporting DETAILS</td>
</tr>
<tr>
<td></td>
<td>Understanding SYNTACTIC structure of sentence and clause</td>
<td>Making INFERENCES</td>
</tr>
</tbody>
</table>

Khalifa and Weir (2009)
Cognitive processes in practice

Questions 1 and 2

Local level

Careful reading

- Understanding LEXICAL and GRAMMATICAL meaning
- Understanding SYNTACTIC structure of sentence and clause

Global level

- Understanding MAIN IDEAS and supporting DETAILS
- Making INFERENCES

Question 3

Khalifa and Weir (2009)
Are the test tasks appropriate for the test construct?
Task types

• No good/bad task type: **FITNESS-FOR-PURPOSE** of tasks and tests

• **RANGE OF TASK TYPES** in a test reduces problems connected with certain task types
Advantages:
• easy to mark
• questions can tap into different levels of cognitive processing

Limitations:
• high chance of getting correct answer through guessing
• task is not authentic

For example: multiple-choice task

1. How many children went on the school trip?

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>12</td>
<td>18</td>
</tr>
</tbody>
</table>

Man: Did you enjoy your school trip, James? Did everyone in your class go?

James: No, twelve people stayed behind, so there were eighteen of us. We had a great time.

Man: Where did you all sleep?

James: We had small rooms with six beds in each one. It was lots of fun.

Now listen again.
For example: paired speaking task

Advantages:
• high authenticity
• interactional skills

Limitations
• role of personality, age, gender, language ability, etc.
• scoring reliability
• practicality
Task types

- No good/bad task type ➔ fitness-for-purpose
- Range of task types
Reliability

How far can we DEPEND on the scores from the test?

Are the test scores CONSISTENT and ACCURATE?
Grade this apple on a scale of 1 (lowest) to 6 (highest).
Grade this apple on a scale of 1 to 6 for **quality of colour**.
How can reliability be increased?

- Clear item/task **INSTRUCTIONS**
- Clear assessment **SCALES** and **CRITERIA**
- **TRAINED** examiners/teachers to deliver tests + score tests
## Rating scales

*Cambridge English: First* Speaking Examiners use a more detailed version of the following assessment scales, extracted from the overall Speaking scales on page 41:

<table>
<thead>
<tr>
<th>Band</th>
<th>Grammar and Vocabulary</th>
<th>Discourse Management</th>
<th>Pronunciation</th>
<th>Interactive Communication</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>5</strong></td>
<td>Shows a good degree of control of a range of simple and some complex grammatical forms. Uses a range of appropriate vocabulary to give and exchange views on a wide range of familiar topics.</td>
<td>Produces extended stretches of language with very little hesitation. Contributions are relevant and there is a clear organisation of ideas. Uses a range of cohesive devices and discourse markers.</td>
<td>Is intelligible. Intonation is appropriate. Sentence and word stress is accurately placed. Individual sounds are articulated clearly.</td>
<td>Initiates and responds appropriately, linking contributions to those of other speakers. Maintains and develops the interaction and negotiates towards an outcome.</td>
</tr>
<tr>
<td><strong>4</strong></td>
<td>Performance shares features of Bands 3 and 5.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>3</strong></td>
<td>Shows a good degree of control of simple grammatical forms, and attempts some complex grammatical forms. Uses a range of appropriate vocabulary to give and exchange views on a range of familiar topics.</td>
<td>Produces extended stretches of language despite some hesitation. Contributions are relevant and there is very little repetition. Uses a range of cohesive devices.</td>
<td>Is intelligible. Intonation is generally appropriate. Sentence and word stress is generally accurately placed. Individual sounds are generally articulated clearly.</td>
<td>Initiates and responds appropriately. Maintains and develops the interaction and negotiates towards an outcome with very little support.</td>
</tr>
<tr>
<td><strong>2</strong></td>
<td>Performance shares features of Bands 1 and 3.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>1</strong></td>
<td>Shows a good degree of control of simple grammatical forms. Uses a range of appropriate vocabulary when talking about everyday situations.</td>
<td>Produces responses which are extended beyond short phrases, despite hesitation. Contributions are mostly relevant, despite some repetition. Uses basic cohesive devices.</td>
<td>Is mostly intelligible, and has some control of phonological features at both utterance and word levels.</td>
<td>Initiates and responds appropriately. Keeps the interaction going with very little prompting and support.</td>
</tr>
<tr>
<td><strong>0</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Performance below Band 1.
Test impact

Key concept

What is the effect of the test on learning and on education systems?
Increasing positive impact

- Test those abilities you want learners to develop and not what is easiest to test

**Test:**
- reading aloud
- describing pictures
- talking about a topic

**Classroom:**
- discussing in pairs/small groups
- describing visuals
- asking and answering questions
- information gap activities
- reading texts aloud (pronunciation practice)
- completing a dialogue
- presentations
Good tests are those which:
• test what they set out to test
  = have **CONSTRUCT VALIDITY**
• produce scores which can be trusted
  = have **RELIABILITY**
• support learning
  = have **POSITIVE IMPACT**

... and also are **PRACTICAL** to develop and deliver
If you’d like to find out more...
Useful tools for you

- EnglishProfile
- ALTE
- Cambridge English
  Write & Improve (BETA)
- Manual for Language Test Development and Examining

© Cambridge English Language Assessment 2016
Teaching English

Free resources for teachers

Sample papers, handbooks, lesson plans and teacher guides

Teaching Qualifications

For new and experienced teachers

Teaching Framework

Helps you assess which stage you are at in your professional development and work out where you want to get to next

Webinars

About our exams and teaching

www.cambridgeenglish.org/teaching-english

© Cambridge English Language Assessment 2016
Examining English - the complete set

www.cambridgeenglish.org/silt
www.cambridge.org/elt/silt