

# **EALTA's Guidelines for Good Practice: a test of implementation**

J. Charles Alderson  
Lancaster University

Jayanti Banerjee  
University of Michigan

## **1. Outline**

1. Background
  - a. The EALTA Guidelines for Good Practice in Language Testing and Assessment
  - b. The 'ELPAC' Validation Study
2. Applying the EALTA Guidelines
3. Issues arising
4. Taking the EALTA Guidelines forward

## **2. EALTA Guidelines for Good Practice (Section C)**

### ***Considerations for test development in national or institutional testing units or centres***

EALTA members involved in test development will clarify to themselves and appropriate stakeholders (teachers, students, the general public), and provide answers to the questions listed under the headings below. Furthermore, test developers are encouraged to engage in dialogue with decision makers in their institutions and ministries to ensure that decision makers are aware of both good and bad practice, in order to enhance the quality of assessment systems and practices.

#### **1. TEST PURPOSE AND SPECIFICATION**

1. How clearly is/are test purpose(s) specified?
2. How is potential test misuse addressed?
3. Are all stakeholders specifically identified?
4. Are there test specifications?
5. Are the specifications for the various audiences differentiated?
6. Is there a description of the test taker?
7. Are the constructs intended to underlie the test/subtest(s) specified?
8. Are test methods/tasks described and exemplified?
9. Is the range of student performances described and exemplified?
10. Are marking schemes/rating criteria described?
11. Is test level specified in CEFR terms? What evidence is provided to support this claim?

#### **2. TEST DESIGN and ITEM WRITING**

1. Do test developers and item writers have relevant experience of teaching at the level the assessment is aimed at?
2. What training do test developers and item writers have?
3. Are there guidelines for test design and item writing?
4. Are there systematic procedures for review, revision and editing of items and tasks to ensure that they match the test specifications and comply with item writer guidelines?
5. What feedback do item writers receive on their work?

### **3. QUALITY CONTROL and TEST ANALYSES**

1. What quality control procedures are applied?
2. Are the tests piloted?
3. What is the normal size of the pilot sample, and how does it compare with the test population?
4. What information is collected during piloting? (teachers' opinions, students' opinions, results,...)
5. How is pilot data analysed?
6. How are changes to the test agreed upon after the analyses of the evidence collected in the pilot?
7. If there are different versions of the test (e.g., year by year) how is the equivalence verified?
8. Are markers trained for each test administration?
9. Are benchmarked performances used in the training?
10. Is there routine double marking for subjectively marked tests? Is inter and intrarater reliability calculated?
11. Is the marking routinely monitored?
12. What statistical analyses are used?
13. What results are reported? How? To whom?
14. What processes are in place for test takers to make complaints or seek reassessments?

### **4. TEST ADMINISTRATION**

1. What are the security arrangements?
2. Are test administrators trained?
3. Is the test administration monitored?
4. Is there an examiner's report each year or each administration?

### **5. REVIEW**

1. How often are the tests reviewed and revised?
2. Are validation studies conducted?
3. What procedures are in place to ensure that the test keeps pace with changes in the curriculum?

### **6. WASHBACK**

1. Is the test intended to initiate change(s) in the current practice?
2. What is the washback effect? What studies have been conducted?
3. Are there preparatory materials?
4. Are teachers trained to prepare their students for the test/exam?

### **7. LINKAGE TO THE COMMON EUROPEAN FRAMEWORK**

1. What evidence is there of the quality of the process followed to link tests and examinations to the Common European Framework?
2. Have the procedures recommended in the Manual and the Reference Supplement been applied appropriately?
3. Is there a publicly available report on the linking process?

### 3. Changes to EALTA GGP

1. Glossing terms:
  - a. Specifications = test blueprint
  - b. Stakeholders = those who commissioned the test or have an interest in it
  - c. Construct = the skills and sub-skills that the test / subtest(s) are intended to measure
2. Deleting questions:
  - a. *What quality control procedures are applied?* (redundant)
  - b. *How is potential test misuse addressed?* (highly unlikely?)
3. Specifying  
*E.G. GGP Q 1.5 Are the specifications for the various audiences differentiated?*

*AET Q 1.4 Which of the following groups of test-takers are specified?*

- i. *Air Traffic Controllers*
- ii. *Pilots*
- iii. *Air Service Personnel*

*Please add any groups your test caters for but which we have not listed.*

4. Clarifying  
*E.G. GGP 4.1 What are the security arrangements?*

*AET Q 5.1 How is cheating (for example, personation, bribery, copying, access to illicit materials, etc) checked or prevented?*

5. Making relevant  
*E.G. GGP Q 5.3 What procedures are in place to ensure that the test keeps pace with changes in the curriculum?*

*AET 6.2 What procedures are in place to ensure that the test keeps pace with changes in the ICAO requirements?*

6. Adding items
  - a. Needs analysis
  - b. The GGP makes no reference to pass/fail distinctions or grade boundaries, whereas in the case of the Aviation English tests, pass/fail boundaries are clearly crucial:  
*AET Q3.9 How is the pass mark for objectively scored tests determined?*
  - c. *AET Q 5.5: What processes are in place for test-takers to make complaints or seek scrutiny of marks, and/or re-marking of the test?*
  - d. *AET Q 5.6: How often can a candidate re-take the test?*

7. Modifying double barrelled questions  
*E.G. GGP Q 1.10 Are marking schemes/rating criteria described?*

*AET Q 1.8 Are the rating criteria available?*

8. Changing ambiguous wording  
*E.G. GGP Q 3.2 Are the tests piloted?*

*AET Q 4.1 Are the tests trialled?*

## 4. Requests for evidence of claims

1. The statement of purpose for your test(s)
2. The construct description
3. Lists or examples of test methods and tasks, or sample practice tests provided for test preparation
4. Examples of test-taker performance
5. The rating criteria
6. The evidence to justify the ICAO level of your test (s)
7. The different kinds of experience demanded of test designers
8. The nature of the training the test developers have received.
9. The different kinds of experience demanded of item writers.
10. The nature of the training the item writers have received.
11. Guidelines for test design and item writing.
12. Documentation of systematic procedures for review and editing of items and tasks
13. Examples of the feedback given to item writers
14. A report of any investigation of test-takers' needs
15. Any survey of how language is used in the AET context
16. The software training procedures used for computer-based scoring.
17. The measures in place to monitor the performance of the computer scoring
18. The training provided to raters of subjectively marked tests.
19. Details of how exemplar performances (benchmarks) are selected.
20. Details of how double-marking is organised
21. Details of the action taken in the event of disagreement between raters
22. Details of the system adopted to monitor the quality of the rating
23. Any report on how the pass mark is determined for objectively scored tests
24. Any report of how the tests are trialled.
25. Evidence or reports of:
  - a. test-takers' views on the difficulty and appropriacy of the test and tasks
  - b. invigilators' views on the difficulty and appropriacy of the test and tasks
  - c. examiners' views on the difficulty and appropriacy of the test and tasks
26. A report of the analysis of trial data.
27. An account of how changes to the test are agreed upon after the analyses of the evidence collected in the trial
28. A report of how the equivalence is verified of different versions of the test.
29. Any report of what statistical analyses are used to examine the results of the tests
30. Any report on how cheating is checked or prevented.
31. Details of how test invigilators are trained
32. Details of how test administration is monitored
33. An example of an invigilator's report.
34. The complaints procedures and how requests are made to seek scrutiny of marks, and/or re-marking of the test