

## **EALTA GUIDELINES FOR GOOD PRACTICE IN LANGUAGE TESTING AND ASSESSMENT**

The EALTA Executive Committee appointed a Working Group to review and revise the draft Code of Practice received from the ENLTA Project. The task of the Working Group was to develop a code of practice or set of guidelines for good practice in testing and assessment which is appropriate to EALTA and its mission.

The EALTA Mission Statement is as follows:

The purpose of EALTA is to promote the understanding of theoretical principles of language testing and assessment, and the improvement and sharing of testing and assessment practices throughout Europe.

The rationale for the Guidelines adopted by the Working Group was the following:

Europe is a multilingual continent, where the diversity of languages, cultures and traditions is highly valued. Part of such diversity is diversity in education systems and assessment traditions and values.

Consequently, EALTA members will strive to adhere to the principles of transparency, accountability and quality appropriate to their particular contexts and spheres of professional involvement in language testing and assessment.

Reflecting its policy of inclusiveness, EALTA wishes to serve the needs of a very broad membership. EALTA's guidelines for good practice in testing and assessment are accordingly addressed primarily to three different audiences: those involved in (a) the training of teachers in testing and assessment, (b) classroom testing and assessment, and (c) the development of tests in national or institutional testing units or centres.

For all these groups, a number of general principles apply: respect for the students/examinees, responsibility, fairness, reliability, validity and collaboration among the parties involved. These general principles are laid out in relevant existing codes of practice, which EALTA members are encouraged to consult in order to further inform the professionalism and quality of their work. Please refer to the following links:

<http://www.iltaonline.com>

<http://www.alte.org>

<http://www.qca.org.uk>

<http://www.apa.org/science/standards.html>

<http://www.apa.org/science/FinalCode.pdf>

EALTA's own guidelines to good practice in language testing and assessment are as follows:

## **A. Considerations for teacher pre-service and in-service training in testing and assessment**

EALTA members involved in teacher training related to testing and assessment will clarify to themselves and appropriate stakeholders (trainees, practising teachers, curriculum developers):

1. How relevant is the training to the assessment context of the trainees?
2. How aware are trainees made of the range of assessment procedures appropriate to their present or future needs?
3. How clearly are the principles of testing and assessment (e.g. validity, reliability, fairness, washback) related to the trainees' context?
4. What is the balance between theory and practice in the training?
5. How far are the trainees involved in developing, trialling and evaluating assessment procedures?
6. How far are trainees involved in marking or assessing student performances?
7. What attention is given to the appropriate analysis of assessment results?
8. What account is taken of trainees' views on the appropriacy and accuracy of assessment procedures?
9. How far do assessment procedures used to evaluate the trainees follow the principles they have been taught?

## **B. Considerations for classroom testing and assessment**

EALTA members involved in classroom testing and assessment will clarify to themselves and appropriate stakeholders (especially pupils/students and as far as possible parents):

### **1. ASSESSMENT PURPOSE(S) AND SPECIFICATION**

1. What is the purpose of the assessment?
2. How does the assessment purpose relate to the curriculum?
3. Are there any test specifications?
4. How well is the curriculum covered?
5. How are the assessment purposes and specifications made known and discussed?

### **2. ASSESSMENT PROCEDURES**

1. Who designs the assessment procedures?
2. How appropriate are the assessment procedures to the learners?
3. How is information on students' learning collected?
4. How is information on students' learning assessed and stored?
5. What efforts are made to ensure that the assessment results are accurate and fair?
6. What efforts are made to promote agreement in marking practices across teachers and schools?
7. What account is taken of students' views on the assessment procedures?

### **3. CONSEQUENCES**

1. What use is made of the results?
2. What action(s) will be taken to improve learning?
3. What kind of feedback do students get?
4. What processes are in place for students or their guardians to make complaints or seek re-assessments?
5. What are the consequences of the assessment procedures for classroom practices?
6. What are the consequences of the results of the assessment for learners?

## **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 intra-rater 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 re-assessments?

#### **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?

#### **D. LINKAGE TO THE COMMON EUROPEAN FRAMEWORK – CEFR**

Linking to the CEFR is a complex endeavour, which may often take the form of a project to be developed along a number of years. Linking exams or tests to a standard such as the CEFR requires a scientific approach, and claims must be based on the results of research, preferably submitted for peer review.

Institutions/exam providers wishing to claim linkage to the CEFR are accountable for the provision of sufficient convincing evidence for such linkage. The following considerations may be useful to gather such evidence.

EALTA members involved in the process of linking an exam to the CEFR will clarify to themselves and appropriate stakeholders (teachers, students, policy makers, the general public):

1. Have issues of test quality and impact been fully addressed? (see EALTA Guidelines Section C). Is information on the quality of the test publicly available?
2. Are those involved in test development all familiar with the CEFR? Have familiarisation with the CEFR procedures been followed? What evidence is there of the procedures followed and the corresponding results?
3. Have the test content and the test specifications been analysed in relation to the CEFR descriptors? What evidence is there of the procedures followed and the corresponding results?
4. Have standardisation procedures been completed for written and spoken performances? Which ones? By whom? What evidence is there of the procedures followed and the corresponding results?

5. Are local performance samples benchmarked to the CEFR publicly available?
6. Have standardisation procedures been completed for test items targeting receptive skills? Which ones? By whom? What evidence is there of the procedures followed and the corresponding results?
7. Are local CEFR-linked items and tasks publicly available?
8. What standard setting procedures have been used to establish cut-off scores for the relevant CEFR level(s)? How many judges have been involved? Is there a standard setting report publicly available?
9. Has validity evidence been collected covering the process of linkage? Is it publicly available?
10. Is there a scheme of level setting that guarantees quality standards and linkage to the CEFR? Is it publicly available?