ALTE Quality Assurance Checklists

Unit 1

Test Construction

Name(s) of people completing this checklist:

Which examination are the checklists being completed for?

At which ALTE Level is the examination at?

Date of completion:

ALTE Quality Assurance Checklist
Unit 1: Test Construction

Instructions for completing checklists.

1.   On each sheet you will see various columns. The first one is headed 'Code of Practice' and each page includes one or sometimes two question(s) or point(s) under that column. These are about the main points from the Code of Practice.

2.   In the second column are Focus Points, asking for more detailed information about the question under the Code of Practice.

3.   For each focus point, please do the following:

a.   If the question can be answered by Yes or No, put a cross in the appropriate column.
b.   Under 'Comments and Reference' add in further **short** information. This might be a reference to certain documents or as an explanation of why you have ticked Yes or No.
c.   In the final column, headed 'Self Evaluation', you will see four boxes for each Focus Point. These are headed as follows:

        IMP = In need of improvement
        UR = Under review
        AD = Adequate
        GP = Good practice

        For each Focus Point you should tick one of these boxes, depending on whether i**n your opinion** this needs to be improved within your organisation (IMP), whether this porcess is being reviewed by your organisation (UR), is adequately dealt with in your organisation (AD), or is good practice (GP).

4.   At the end of the Unit you will find questions from the Code of Practice column repeated in Schedule A. Here you can add any longer information there was not room for in the boxes.

Example of a completed checklist – this is to give an example of how much information should be added to this part of the checklists.
Please add longer comments in Schedule A at the end of the Unit.

Unit 1 - Test Construction								
Code of Practice	**Focus Points**			**Explanations and / or references**	**Self evaluation**			
		Yes	No		INP	UR	AD	GP
A. Conceptual phase. *Design and development* A1. Describe how and why the exam was developed.	i. Who or what organisation decided that the examination was required? **For example**: was an external body such as the Ministry of Education involved?			National Education Act 1998 provides guidelines for examinations of language as a foreign language.				
	ii. If yes, what influence did they have over the design and development?			Final examination papers to be approved by committee set up under the Act, otherwise ALTE member responsible for all stages of process. (See examination handbook introduction page 10)				
	iii. If no, what other factors influenced the design and development? **For example**: historical developments over time: changes in curriculum or technology?							

Please start completing the checklists from here

Unit 1 - Test Construction								
Code of Practice	**Focus Points**			**Explanations and / or references**	**Self evaluation**			
		Yes	No		INP	UR	AD	GP
A. Conceptual Phase - *Design and development* A1. Describe how and why the exam was developed.	i. Who or what organisation decided that the examination was required?							
	ii. Was an external body involved, for example, the Ministry of Education?							
	ii. If yes, what influence did they have over the design and development?							
	iii. If no, what other factors influenced the design and development? For example: historical developments over time, changes in **curriculum** or technology?							

Unit 1 - Test Construction									
Code of Practice	**Focus Points**			**Explanations and / or references**		**Self evaluation**			
		Yes	No			INP	UR	AD	GP
Purpose and context of use. A2. Describe what the examination assesses	i. Do you test the following skills in the examination?								
	Reading?								
	Listening?								
	Writing?								
	Speaking?								
	ii. If yes, how are these **constructs** defined?								
	iii. Do you relate your examination to a model (or models) of communication and **communicative language ability or competence**?								
	iv. How do you cover features such as overall **proficiency**, grammar and vocabulary?.								

Unit 1 - Test Construction								
Code of Practice	**Focus Points**			**Explanations and / or references**	**Self evaluation**			
		Yes	No		INP	UR	AD	GP
A3. Describe what the examination should be used for.	i. Do you describe what the examination should be used for? For example: as an end of course **assessment**, as part of an educational entry requirement, as part of a procedure for citizenship, as a job entry requirement, as part of procedures for immigration?							
	ii. Do you intend the examination to be used for national purposes (within your country / region) or internationally, or both?							
	a) National							
	b) International							
	c) Both							

Unit 1 - Test Construction								
Code of Practice	**Focus Points**			**Explanations and / or references**	**Self evaluation**			
		Yes	No		INP	UR	AD	GP
A4. Describe the population for which the population is appropriate.	I. Do you describe which groups of candidates the examination is suitable for? **For example:** Students	▓	▓					
	Young children							
	Asylum seekers							
	Immigrants							
	Adult learners: for professional reasons / for reasons of tourism or social contact / for academic reasons.							
	ii. How was the population established in terms of the purpose of the examination?	▓	▓					
	iii. Do you verify that the population taking the exam is the same as the intended one? If so, how?							

Unit 1 - Test Construction								
Code of Practice	**Focus Points**			**Explanations and / or references**	**Self evaluation**			
		Yes	No		INP	UR	AD	GP
B. Test development, test construct and context. Routine procedures. B1. Explain how the content and skills to be tested are selected (for each component of the examination?	i. Were you required to make reference to an official source in establishing the content of the examination, such as national curriculum guidelines? **For example**: from a Ministry?							
	ii. Do you make use of domain or profile descriptions (skills and components?)							
	iii. Is any use made of the Common European Framework of Reference? If yes, how?							
	iv. Were the specific assessment goals developed by your examination body itself? If so, how?							

Unit 1 - Test Construction								
Code of Practice	**Focus Points**			**Explanations and / or references**	**Self evaluation**			
		Yes	No		INP	UR	AD	GP
B2. Describe the routine procedures used to ensure the appropriateness of the examination for the intended candidates.	In producing test materials do you take into account the following features of the intended candidates?							
	Racial background							
	Ethnic background							
	Linguistic background (L1)							
	Language learning background							
	Age							
	Socio-economic background							
	Gender							
	Social-cultural factors							
	Educational level							
	ii. If yes, how do you do this?							
B3. Describe how examination tasks are reviewed and revised to avoid potentially insensitive content or language.	i. Do you use routine procedures to review and revise examination materials to avoid potentially insensitive content or language?							
	ii. If yes, how do you do this? For example, using taboo checklists and review procedures?							

Unit 1 - Test Construction								
Code of Practice	**Focus Points**			**Explanations and / or references**	**Self evaluation**			
		Yes	No		INP	UR	AD	GP
B4. Describe procedures that help to ensure that differences in performance are related to the skills under assessment rather than irrelevant factors.	i. Do you make use of expert judgement such as feedback from external stakeholder groups or reviews by internal expert groups? **For example**: through developing and using 'taboo' lists?							
	ii. Do you make use of **pre-testing**? For example: procedures based of data collection and statistical means to investigate bias?							
	iii. What procedures do you use in constructing items to minimize **bias** by factors such as gender and L1?							

Unit 1 - Test Construction								
Code of Practice	**Focus Points**			**Explanations and / or references**	**Self evaluation**			
		Yes	No		INP	UR	AD	GP
Examination development B5. Describe the process of examination development.	I. Which staff in your organisation are directly involved in examination development? For example: in designing the test, developing specifications and writing the materials?							
	ii. How are they recruited? Do you provide training? If yes, how?							
	iii. What documentation is available to ensure that specifications for the examination are developed effectively? For example, do you use item writer checklists and guidelines?							
	iv. What routine procedures are used for: Task / item selection? Task / item production? Pretesting of materials? Revision of task / items?							

Unit 1 - Test Construction								
Code of Practice	**Focus Points**			**Explanations and / or references**	**Self evaluation**			
		Yes	No		INP	UR	AD	GP
B6. Describe procedures to determine reliability. **For example**: by skill / paper, overall for the whole examination.	i. Are you able to estimate the **reliability** of your tests on a routine basis?							
	ii. If so, how? For example: through data collection and psychometric procedures, through procedures involving expert judgement.							

Unit 1 - Test Construction								
Code of Practice	**Focus Points**			**Explanations and / or references**	**Self evaluation**			
		Yes	No		INP	UR	AD	GP
B7. Describe procedures to determine **validity**. For example: Construct validity, content validity, criterion - related validity	i. Do you have a theoretical basis for defining the constructs underlying the skills tested? For example: a model of language ability or communicative language use							
	ii. Do you collect data and use psychometric procedures to validate these constructs?							
	iii. Do you use procedures to define and control test content? For example: content / domain specifications and content checklists							

Unit 1 - Test Construction								
Code of Practice	**Focus Points**			**Explanations and / or references**	**Self evaluation**			
		Yes	No		INP	UR	AD	GP
C. Communication with external stakeholders. C1. Describe how you provide examination users and takers with information to help them judge whether a particular examination should be taken.	i. In communicating with external stakeholders, do you provide the following documents:							
	Introductory leaflets or flyers?							
	Short brochures, for example with level descriptions for Common European Framework of Reference or ALTE levels?							
	Longer brochures with descriptions of content, target group etc.?							
	In depth handbooks for teachers, including sample materials?							
	ii. How do you ensure that the documentation is suitable for the different user groups? For example: the primary users - candidates; the secondary users - parents, teachers, employers etc.							

Unit 1 - Test Construction								
Code of Practice	**Focus Points**			**Explanations and / or references**	**Self evaluation**			
		Yes	No		INP	UR	AD	GP
C2. Describe how you make such information equally available to all candidates.	i. How do you produce and distribute the information?							
	As printed books, brochures, etc.							
	Electronically on a website							
	Posters / adverts in the press							
	ii. How often is this information revised so that it is kept up to date?							
	iii. Is revised information automatically sent to users - centres, candidates etc?							
	iv. If not, how can users obtain up-to-date information?							
	v. Is information provided free of charge?							
	vi. How do you ensure that the quality of the materials is suitable for the different user groups?							

Unit 1 - Test Construction								
Code of Practice	**Focus Points**			**Explanations and / or references**	**Self evaluation**			
		Yes	No		INP	UR	AD	GP
C3. Describe how you provide candidates with information they need in order to be familiar with coverage of the examination, the type of task formats, the **rubrics** and other instructions and other instructions as appropriate examination-taking	i. Do you provide the following:							
	Complete practice examinations							
	Sample materials by skill							
	Videos for speaking practice test							
	Information about assessment criteria. For example: user oriented assessment criteria							
C4. Describe which samples or complete copies of relevant documents you provide.	Examples of answer sheets							
	Examples of how results will be reported to users							
	Score reports							
	Profiles							
	Certificates							
	ii. Do the users have to pay for some documents and if so, which ones?							
C5. Describe how and for whom you provide them	i. How are these materials distributed to ensure adequate availability?							

Schedule A

Please add below any further information you have in answer to the questions:

A. Conceptual phase.

A1. Describe how and why the examination was developed.

A2. Describe what the examination assesses.

A3. Describe what the examination should be used for.

A4. Describe the population for which the examination is appropriate.

B. Test development, test construct and context.

B1. Describe how the content and skills to be tested are selected (for each component of the examination)

B2. Describe the routine procedures used to ensure the appropriateness of the examination for the intended candidates.

B3. Describe how examination tasks are reviewed and revised to avoid potentially insensitive content or language.

B4. Describe procedures that help to ensure that differences in performance are related to the skills under assessment rather than irrelevant factors.

B5. Describe the process of examination development.

B6. Describe procedures to determine reliability. For example: by skill / paper, overall for the whole examination.

B7. Describe procedures to determine validity. For example: Construct validity, Content validity, Criterion-related validity.

C. Communication with External Stakeholders.

C1. Describe how you provide examination users and takers with information to help them judge whether a particular examination should be taken.

C2. Describe how you make such information equally available to *all* candidates.

C3. Describe how you provide candidates with information they need in order to be familiar with coverage of the examination, the type of task formats, the rubrics and other instructions and appropriate examination-taking strategies.

C4. Describe which samples or complete copies of relevant documents you provide.

C5. Describe how and for whom you provide them.

Term	Definition
Assessment	In language testing, the measurement of one or more aspects of language proficiency, by means of some for of test or procedure.
Bias	A test or item can be considered to be biased if one particular section of the candidates population is disadvantaged by some particular aspect of the test or item which is not relevant to what is being measured. Sources of bias may be connected with gender,
Calibrate	In item response theory, to estimate the difficulty of a set of test items
Certificates	A document stating that a names person has taken a test or component of a test and had achieved a particular grade, usually at least a pass. See also *Diploma*
Clerical markers	A method of marking in which markers do not need to exercise any special expertise or subjective judgement. The mark by following a mark scheme which specifies all acceptable responses to each test item.
Communicative competence	The ability to use language appropriately in a variety of situations and settings.
Construct	A hypothesized ability or mental trait which cannot necessarily be directly observed or measured, for example, in language testing, listening ability. Language tests attempt to measure different constructs which underlie language ability. In addition to language ability itself, motivation, attitude and acculturation are all relevant constructs.
Construct validity	A test is said to have construct validity of the scores can be shown to reflect the theory about the nature of a construct or its relation to other constructs. It could be predicted, for example, that two valid tests of listening comprehension would rank learners in the same way, but each would have a weaker relationship with the scores on a test of grammatical competence.
Constructed response	A form of written response to a test item that involves active production, rather than just choosing from a number of options.

Content validity	A test is said to have content validity if the items or tasks of which it is made up constitute a representative sample of items for the area of knowledge or ability to be tested. These are often related to a syllabus or course.
Co-ordination session	For the assessment of Speaking and Writing human markers (raters / examiners)are required. Six aspects of the process of ensuring that the markers can mark in a reliable and valid way can be identified: RITCME - **R**ecruitment; **I**nduction and Initial Training; **T**raining for the specific exam; **C**o-ordination (before each exam takes place or at least regularly); **M**onitoring of their conduct; **E**valuation of their conduct. A co-ordination session is the event to ensure that all examiners have been co-ordinated effectively before they examine.
Criterion-related validity	A test is said to have criterion-related validity if a relationship can be established between test scores and some external criterion which is believed to be a measure of the same ability. Information on criterion-relatedness is also used in determining how well a test predicts future behaviour.
Cronbach's alpha	A reliability estimate, measuring the internal consistency of a test. It ranges in value from 0 to 1. It is often used for tests with rating scales as opposed to tests with dichotomous items, although it may be used for both. Also referred to as coefficient alpha.
Curriculum	An overall description of the aims, content, organisation, methods and evaluation of an educational course.
Cut score	The minimum score a candidate has to achieve in order to get a given grade in a test or examination. In mastery testing, the score on a test which is considered to be the level required in order to be considered minimally competent or at 'mastery' level.
Difficulty (index)	In classical test theory, the difficulty of an item is the proportion (p) of candidates responding to it correctly. This means that the difficulty estimate of an item is sample dependent, and changes according the level of ability of candidates.

Diploma	A document stating that a names person has taken a test or component of a test and had achieved a particular grade, usually at least a pass. Often interpreted as being of a higher level qualification than a certificate. See also *Certificate*
Examiner	Someone who assigns a score to a candidate's responses in a test, using subjective judgement to do so. Examiners are usually qualified in the relevant field and are required to undergo a process of training and standardization. In oral testing the roles of examiner and interlocutor are sometimes distinguished. Also referred to as assessor or rater.
Grading	The process of converting test scores or marks into grades.
Impact	The effect created by a test, both in terms of influence on general education process, and in terms of the individuals who are affected by the results.
Internal consistency (sample / estimate)	A feature of a test, represented by the degree to which candidates' scores on individual items in a test are consistent with their total score. Estimates of internal consistent can be used as indices of test reliability, various indices can be computed, for example KR-20 alpha. See also *Cronbach's alpha*
Invigilator	A person of authority employed at an examination centre to ensure that the exam is conducted according to the established procedures.
Marker	Someone who assigns a score to a candidate's responses to a written test. This may involve the use of expert judgement, or in the case of a clerical marker, the relatively unskilled application of a mark scheme.
Marking	Assigning a mark to a candidate's responses to a test. This may involve professional judgement, or application of a mark scheme which lists all acceptable responses.
Optical mark reader (OMR)	An electronic device used for reading information directly from answer sheets or mark sheets. Candidates or examiners can mark item responses or tasks on a mark sheet and this information can be read directly into a computer. Also referred to as scanner.

Performance	The act of producing language by speaking or writing. Performance, in terms of language actually produced by people, is often contrasted with competence, which is the underlying knowledge of a language.
Population sample	A selection of a sub-set of elements from a population.
Pretesting	A stage in the development of test materials at which items are tried out with representative samples from the target population in order to determine their difficulty. Following statistical analysis, those items that are considered to be satisfactory can be used in live tests.
Proficiency	Knowledge of a language and a degree of skill in using it.
Rater	See definition for examiner
Rater agreement	The degree of agreement between two assessments of the same sample of performance made at different times by the same assessor. This has particular relevance to the assessment of speaking and writing skills in tests where subjective judgements by examiners are required.
Regulations	An official document provided by the examination board which states the conditions under which enrolment for the exams, the conduct of the exams and the issue of results will be made. Candidates need to be aware of the regulations before the take the exam, including the rights and obligations they are signing up to.
Reliability	The consistency or stability of measures from a test. The more reliable a test is, the less random error it contains. A test which contains systematic error, e.g. bias against a certain group, may be reliable, but not valid. See also *Test - Retest*
Results	The outcome of a test, as reported to a test taker or test user.
Rubrics	The instructions given to candidates to guide their responses to a particular test task.
Score	A) The total number of points someone achieves in a test, either before scaling (raw score) or after scaling (scaled score). B) To assign numerical values to observed performance.

Standard error of measurement (SEM)	In classical true score test theory, the SEM is an indication of the imprecision of a measurement. The size of the standard error of measurement depends of the reliability and the standard deviation of the test scores.
Standardisation	The process of ensuring that assessor adhere to an agreed procedure and applying rating scales in an appropriate way.
Supervisor	A senior invigilator who is responsible for the conduct of an examination at an examination centre or in the examination room.
Test-retest	An estimate of reliability obtained by administering the same test to the same candidates in the same conditions, and correlating the scores on two sittings. It is concerned with the stability of scores over time, and is also appropriately used where estimates of internal consistency are not possible.
Validity	The extent to which scores on a test enable inferences to be made which are appropriate, meaningful and useful, given the purpose of a test. Different aspects of validity are identified, such as content, criterion and construct validity; these provide different kinds of evidence for judging the overall validity of a test for a given purpose. See also: *Construct validity, content validity, criterion related validity*
Discrimination	The power of an item to discriminate between weaker and stronger candidate. Various indices of discrimination are used. Some (e.g. biserial, point-biserial) are based on the correlation between the score on the item and a criterion, such as the total score on the test or some external measure of proficiency. Others are based on the difference in the item's difficulty for high and low ability groups. In item response theory the 2, and 3, parameter models estimate item

ALTE Quality Assurance Checklist
Unit 1: Test Construction

Question (e.g. C2i)	Answer