



ALTE Quality Assurance Checklists

Unit 3

Marking, Grading & Results

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:

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 **in your opinion** this needs to be improved within your organisation (IMP), whether this process is being reviewed by your organisation (UR), is adequately dealt with in your organisation (AD), or is good

4. At the end of the Unit you will find questions from the Code of Practice column repeated in Schedule C. Here you can add any longer information there was not room for in the boxes.
5. Please complete the document **electronically** and e-mail or send it on disk to the Secretariat by 1 February
6. **At the moment please do not send any supporting documents, only the questionnaire, even if you have referred to other documents in your answers.**

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 C at the end of the Unit.

Unit 3. Marking, Grading & Results								
Code of Practice	Focus Points	Focus Points		Explanations and / or references	Self evaluation			
		Yes	No		INP	UR	AD	GP
G. Marking G1. Describe the procedures used to mark the examination papers.	i. Which procedures do you used to ensure accurate and reliable marking of the examination? For example: Do you use:			<i>Described in the examination handbook.</i>				
	Optical mark readers?	x		<i>For Reading, Listening and Structural Competence papers.</i>		x		
	Clerical markers?	x		<i>For Writing papers</i>		x		
	Professional examiners?	x		<i>For the Speaking test.</i>		x		

Unit 3. Marking, Grading & Results								
Code of Practice	Focus Points			Explanations and / or references	Self evaluation			
		Yes	No		INP	UR	AD	GP
G. Marking G1. Describe the procedures used to mark the examination papers.	i. Which procedures do you used to ensure accurate and reliable marking of the examination? For example: Do you use:							
	Optical mark readers?							
	Clerical markers?							
	Professional examiners?							

Unit 3. Marking, Grading & Results								
Code of Practice	Focus Points	Focus Points		Explanations and / or references	Self evaluation			
		Yes	No		INP	UR	AD	GP
G2. Describe the procedures used to ensure that marking is carried out accurately including recruitment, training and monitoring of the examiners (for example for speaking and writing).	i. For optical or clerical methods what procedures are used to check accuracy?							
	ii. How are new clerical markers and professional examiners selected and trained?							
	iii. What routine training or co-ordination procedures are used for markers and examiners / raters ?							
	iv. Do examiners for speaking / writing attend regular co-ordination session ? If so, is this obligatory and how often does it take place?							
	v. How is this carried out to ensure standardisation ?							

Unit 3. Marking, Grading & Results								
Code of Practice	Focus Points	Focus Points		Explanations and / or references	Self evaluation			
		Yes	No		INP	UR	AD	GP
	vi. Do you use single or double rating methods for speaking and writing components?							
	vii. What types of monitoring of markers and examiners takes place? For example: what checking methods are used to detect possible human error?							
	viii. What procedures are used when differences between raters occur?							

Unit 3. Marking, Grading & Results								
Code of Practice	Focus Points	Focus Points		Explanations and / or references	Self evaluation			
		Yes	No		INP	UR	AD	GP
H. Grading H1. Describe the procedures used to establish pass marks and / or grades and cut scores . How is the standard set and maintained?	i. How do you establish your assessment scales and link them to the ALTE Framework / Common European Framework of Reference?							
	ii. How do you set cut scores for pass/ fail and other boundaries?							
	iii. How are the different grade boundaries decided?							
	iv. How do you make comparisons between parallel examinations in subsequent sessions to ensure stability and equivalence?							

Unit 3. Marking, Grading & Results								
Code of Practice	Focus Points			Explanations and / or references	Self evaluation			
		Yes	No		INP	UR	AD	GP
H2. Describe how you ensure that differences in performance are related primarily to skills under assessment rather than to irrelevant factors.	i. What procedures are used to detect test bias introduced by factors such as L1, country of origin, gender, age and race / ethnic origin?							
	ii. How is this information used when grading the examination?							

Unit 3. Marking, Grading & Results								
Code of Practice	Focus Points	Focus Points		Explanations and / or references	Self evaluation			
		Yes	No		INP	UR	AD	GP
J. Results J1. Describe how results are reported and what information is provided to test users that will help them interpret the results.	i. Do you issue pass / fail results or do you use a single reporting scale with band scores?							
	ii. If you use a single scale, how is this calibrated , for example, using statistical means?							
	iii. How is the stability of the scores / grades maintained over time?							

Unit 3. Marking, Grading & Results								
Code of Practice	Focus Points	Focus Points		Explanations and / or references	Self evaluation			
		Yes	No		INP	UR	AD	GP
	iv. Do you establish a standard error of measurement (SEM)? If so, how is this reported?							
	v. When and how are the results for the examination issued? For example: by mail, by email, on the web etc.							
	vi. How do you relate the examination to the Common European Framework of Reference and the ALTE Framework? For example: do you use Can Do statements?							

Unit 3. Marking, Grading & Results								
Code of Practice	Focus Points	Focus Points		Explanations and / or references	Self evaluation			
		Yes	No		INP	UR	AD	GP
J2. Describe how prompt and easily understood reports of examination results are delivered that describe candidate performance clearly and accurately.	i. What information is provided to assist users in the interpretation of results?							
	ii. In what form are results reported to candidates? For example: as certificates, diplomas , scores on rating scales, profiles etc.							
	iii. If certificates are issued, how is this done?							

Unit 3. Marking, Grading & Results								
Code of Practice	Focus Points	Focus Points		Explanations and / or references	Self evaluation			
		Yes	No		INP	UR	AD	GP
J3. Describe how users are warned to avoid specific, reasonably anticipated misuses of examination results.	i. How do you protect against fraud, such as impersonation and illegal copying of certificates.							
	ii. How do candidates get their certificates? For example: when are they issued and in what format?							

Unit 3. Marking, Grading & Results								
Code of Practice	Focus Points	Focus Points		Explanations and / or references	Self evaluation			
		Yes	No		INP	UR	AD	GP
J4. Describe the rights which candidates may or may not have to obtain copies of papers and completed answer sheets, to re-take papers, have papers re-marked or results checked.	i. Do you provide an explanation to users of the appropriate uses of the examination results? For example: how to interpret the results for the intended purpose or purposes?							
	ii. Do you provide warnings of possible situations for which the use of the examination results would not be suitable?							

Unit 3. Marking, Grading & Results								
Code of Practice	Focus Points	Focus Points		Explanations and / or references	Self evaluation			
		Yes	No		INP	UR	AD	GP
J5. Describe how long information about results will be kept on file and indicate to whom and under what circumstances examination results and associated data / materials will or will no be released.	i. What rights do candidates have in relation to their examination papers and data stored by your institution about them?							
	ii. Are these rights restricted by the 'contract' they sign when they enter for the exam?							
	iii. If so, in what ways?							

Unit 3. Marking, Grading & Results								
Code of Practice	Focus Points	Focus Points		Explanations and / or references	Self evaluation			
		Yes	No		INP	UR	AD	GP
	iv. Which procedures are applied and under what circumstances are the results and associated data / materials released?							
	v. What procedures do you have in place for data protection?							
	vi. Do these procedures conform to national and international regulations?							
	vii. How is this regulated within your organisation to ensure compliance?							

Schedule C

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

G. Marking

- G1.** Describe the procedures used to mark examination papers.
- G2.** Describe the procedures used to ensure that marking is carried out accurately including the recruitment, training and monitoring of the examiners.

H. Grading

- H1.** Describe the procedures used to establish pass marks and / or grades. How is the standard set and maintained.
- H2.** Describe how you ensure that differences in performance are related primarily to the skills under assessment rather than to irrelevant factors.

J. Results

- J1.** Describe how results are reported and what information is provided to test users that will help them interpret the results.
- J2.** Describe how prompt and easily understood reports of examination results are delivered that describe candidate performance clearly and accurately.
- J3.** Describe how users are warned to avoid specific, reasonably anticipated misuses of examination results.
- J4.** Describe the rights which candidates may or may not have to obtain copies of papers and completed answer sheets, to re-take papers, have papers re-marked or results checked.
- J5.** Describe how long information about results will be kept on file and indicate to whom and under what circumstances examination results and associated data / materials will or will not be released.

Term	Definition
Assessment	In language testing, the measurement of one or more aspects of language proficiency, by means of some form 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 named 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 if 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: R ITCME - R ecruitment; I nduction and I nitial T rainig; T rainig 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 to the level of ability of candidates.
Diploma	A document stating that a named 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 consistency can be used as indices of test reliability, various indices can be computed, for example KR-20 alpha. See also <i>Cronbach's alpha</i>
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 <i>Test - Retest</i>
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: <i>Construct validity, content validity, criterion related validity</i>

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

Question (e.g. C2i)	Answer