

Testing and Evaluation Handout I

I Basic Concepts in Testing and Evaluation

a) Kinds of tests and testing

☐ Proficiency Tests	☐ tests people's ability in a language regardless of the learning process that one has gone through
☐ Achievement Test	☐ can be <i>final achievement tests</i> like your final exams ☐ can be <i>progress achievement tests</i> or <i>progress monitored tests</i> : should relate to objectives of the language use & also the way feedback is given
☐ Diagnostic Tests	☐ used to identify learners' strengths and weaknesses ☐ generally used in ESP (English for Specific Purposes)
☐ Placement Tests	☐ intended to provide a certain goal or objective in which the students are supposed to fit in
☐ Norm-referenced testing	☐ is dependent on the test scores of other students e.g. percentile scores ☐ does not tell much about what the student is capable of ☐ most achievement tests are norm-referenced
☐ Criterion-referenced testing	☐ provides information about what the student can perform in the language ☐ the criteria are pre-defined and the students know <i>why</i> they are tested for

II Key Concepts in Testing

☐ Validity	☐ can be of two types – content validity and criterion-related validity ☐ <i>content validity</i> depends on what the test is meant for and whether it is testing that ☐ <i>criterion-related validity</i> relates to examining the test results to highly dependable and independent alternative sources
☐ Reliability	☐ depends on two factors: whether the test can be used repeatedly; and whether the test can be used for various levels of students
☐ Backwash	☐ the effect that tests have on learning and teaching ☐ can be of two types – <i>negative</i> and <i>positive</i>

III Stages of Test Development

☐ Make a clear and full statement of the testing issue ☐ write complete specifications for the test ☐ write and moderate items ☐ trial the items and modify ☐ trial the test on NNS for validity ☐ analyse the results of the trial and make necessary changes ☐ calibrate/ fine tune the criteria ☐ validate the results ☐ handbooks for test takers, test users and staff ☐ training programmes for staff involved

Evaluating Multimedia Materials

Handout II

I Evaluating Multimedia Materials: Question to begin with

Multimedia software	☐ does the software or the activities allow students to interact, modify and negotiate meanings?
Teacher-planned multimedia activities	

II Criteria for Evaluating Multimedia Materials

(adapted from Chapelle, C.A. 2001. *Computer Applications in Second Language Acquisition*. Cambridge: CUP)

Language learning potential	☐ is there a beneficial focus on form? ☐ is there interactional modification? ☐ is there a modification in output? ☐ is there enough support for the learner? ☐ is there control given to the learner?
Learner fit	☐ does the material fit the linguistic level of learner? ☐ does it have a range/variety to suit various levels? ☐ does it provide opportunity to develop?
Meaning Focus	☐ is the focus on meaning and negotiation? ☐ does it involve exchange and accomplishment of goals?
Authenticity	☐ does the material serve the purpose of learning the language? ☐ does it prepare the learner for actual language use?
Positive impact	☐ is the material interesting so that it can develop metacognitive strategies? ☐ does it help in developing pragmatic qualities? ☐ does it have a positive impact on the learner/ teacher/ institution?
Practicality	☐ is the material easy to use under practical conditions? ☐ does the material fit the infrastructure (software and hardware)? ☐ does it fit social and cultural conditions?

III Some other questions to keep in mind:

- ☐ does the material incorporate recent developments in language teaching?
- ☐ does the material include recent developments in computer software?
- ☐ does the material allow students to use interactive platforms like chat or blog (CMC)?
- ☐ does the material involve the students in participatory activities (pair and group work)?
- ☐ any others?

IV Computer Assisted Language Testing (CALT)

If you are evaluating a CALT, then these are some principles to keep in mind:

Reliability	☐ does the test relate learners' performance with language abilities?
Validity	☐ is the test accepted widely and does it reflect the construct?
Authenticity	☐ does the test allow learners to use language authentically as in real life?
Washback	☐ does the test have a positive or a negative washback?
Practicality	☐ is the test cost-effective? ☐ does it take into consideration resources available?

Testing and Evaluation: Scoring Rubrics Handout III

Using scoring rubrics for evaluation of multimedia materials

I A working definition

Rubrics are rating scales-as opposed to checklists-that are used with performance assessments. They are formally defined as scoring guides, consisting of specific pre-established performance criteria, used in evaluating student work on performance assessments.

II Types of rubrics

Holistic rubric	☐ A holistic rubric requires the teacher to score the overall process or product as a whole, without judging the component parts separately
Analytic rubric	☐ the teacher scores separate, individual parts of the product or performance first, then sums the individual scores to obtain a total score

III. Example of a holistic rubric

(Adapted from Mertler, C. 2001. Designing scoring rubrics for your classroom. *Practical Assessment, Research & Evaluation*, 7(25) <http://PAREonline.net/getvn.asp?v=7&n=25>

Template for Holistic Rubrics	
Score	Description
5	Demonstrates complete understanding of the problem. All requirements of task are included in response.
4	Demonstrates considerable understanding of the problem. All requirements of task are included.
3	Demonstrates partial understanding of the problem. Most requirements of task are included.
2	Demonstrates little understanding of the problem. Many requirements of task are missing.
1	Demonstrates no understanding of the problem.
0	No response/task not attempted.

IV. Example of an analytic rubric

Population Sampling					
	Beginning 1	Developing 2	Accomplished 3	Exemplary 4	Score
Sampling Technique	Inappropriate sampling technique used	Appropriate technique used to select sample; major errors in execution	Appropriate technique used to select sample; minor errors in execution	Appropriate technique used to select sample; no errors in procedures	
Survey/ Interview Questions	Inappropriate questions asked to gather needed information	Few pertinent questions asked; data on sample is inadequate	Most pertinent questions asked; data on sample is adequate	All pertinent questions asked; data on sample is complete	
Statistical Analyses	No attempt at summarizing collected data	Attempts analysis of data, but inappropriate procedures	Proper analytical procedures used, but analysis incomplete	All proper analytical procedures used to summarize data	
Communication of Results	Communication of results is incomplete, unorganized, and difficult to follow	Communicates some important information; not organized well enough to support decision	Communicates most of important information; shows support for decision	Communication of results is very thorough; shows insight into how data predicted outcome	