

KEY TERMS OF TESTING AND EVALUATION

Estera STĂNUȘ*

We have considered it useful to offer a selection of basic terms referring to testing and evaluation with definitions and explanations supplied by different authors. The case may be that, in the absence of a single theory of testing and evaluation, some definitions of terms used in a special or technical sense in this field of activity and research vary from one author to another. Nevertheless, the essential characteristics of the realities they describe are to be found in most of them. We have listed the French and the English terminologies separately, leaving our readers the pleasure of comparing them and discovering both the similarities and the different ways of perceiving and interpreting the same reality.

The sources we consulted for compiling this lexical inventory have been indicated at the end.

- **accountability** A requirement of language tests, that they be answerable to the interests and needs of those taking them.
- **achievement tests** Tests which aim to establish what has been learned in a course of instruction.
- **alternative assessment** A movement for the reform of school-based assessment, away from standardized multiple choice tests and towards assessments which are more sensitive to the goals of the curriculum. Typically includes portfolio assessment, exhibitions, records of participation in classroom activities, etc.
- **appraisals** A system of reviewing a person's past, present and future – their work, their problems, their attitudes, etc. – leading from and to negotiated targets.
- **backwash (or washback)** The effect that a final test has on the teaching programme that leads to it – teaching to the test.
- **banding** Description of the test-taker's whole range of ability in the target language. Each band refers to both descriptors and relates to levels of ability. A band describes an ability level in any one

skill, and therefore a writing band would include information concerning relevance, adequacy, grammar etc.

- **cloze test** A test of reading in which a passage is presented with words regularly deleted; the reader's task is to supply the missing words.
- **clinics** Handing over a small part of the syllabus to the students. A section of a lesson (e.g. part of the last class part each month) is dedicated to questions that the students raise, like patients visiting the doctor.
- **competency** A specific practical skill used in the planning and assessment of adult training outcomes.
- **computer adapted test** A test consisting of a bank of items; as items are answered further items are selected which will most contribute to the emerging picture of the candidate's ability. The test thereby adapts itself to each individual candidate.
- **computer based testing (CBT)** Procedures for presenting test material (stimulus and task **prompt**) via computer to candidates, who respond using the computer keyboard.
- **continuous assessment** This is no different from any other form of assessment – it simply refers to how frequently you test or evaluate the students. The student is assessed on not just one performance, but on many performances.
- **content validity** The extent to which the test appropriately samples from the **domain** of knowledge and skills relevant to performance in the **criterion**.
- **criterion** 1. The **domain** of behaviour relevant to test design. 2. An aspect of performance which is evaluated in test scoring, e.g. fluency, accuracy etc.
- **descriptor** One of several description levels of abilities for each sub-skill, e.g. *Complete understanding with appropriate vocabulary and few if any circumlocutions.*
- **direct testing** An approach to testing which

* Junior Assistant Lecturer, Department of Germanic Languages and Business Communication, ASE Bucharest

emphasizes the need for close simulation of the **criterion** in the test. In the testing of speaking, the **Oral Proficiency Interview (OPI)** is a **direct testing** technique.

- **diagnostic tests** These use **proficiency** or **achievement tests** to analyse strengths and weaknesses in the student or the teaching programme itself.
- **face validity** The extent to which a test meets the expectations of those involved in its use, e.g. administrators, teachers, candidates and test score users; the acceptability of a test to its **stakeholders**.
- **formative evaluation** This is the evaluation done and feedback gained during a process so that the process can be changed to make it more effective – to help this year’s students rather than next year’s.
- **high/low correlation** Is the test an extremely good indicator of the skill we are trying to test?
- **indirect testing** Testing things related to the skill we want to test in order to give us an indication of how the student would perform if they did the skill itself. Compare **direct testing**.
- **International English Language Testing System (IELTS)** A performance based test of English for academic purposes, used mainly for the selection of international students at universities in the United Kingdom, Australia and elsewhere.
- **integrative tests** Tests which integrate knowledge of various components of language systems and an ability to produce and interpret language appropriately in context; seen as a necessary complement to discrete point tests.
- **inter-rater reliability** The extent to which pairs of **raters** agree; technically a statistic expressing degree to which the **ratings** of one rater are predictable from the ratings of another, based on the ratings given by each for a given set of test performances
- **Item Response Theory (IRT)** An approach to measurement which uses complex statistical modelling of test performance data to make powerful generalizations about item characteristics, the relation of items to candidate ability, and about the overall quality of tests.
- **job analysis** The work of identifying tasks and roles in the **criterion** setting which can be used as the basis for task design in **performance tests**, particularly in specific purpose contexts.
- **level descriptors** In **rating scales**, statements describing the character of a minimally acceptable performance at a given level.
- **moderation** The process of reconciling or reducing differences in the judgements and standards used by different **raters** within a **rating**

procedure, usually at meetings of raters at which performances at relevant levels are rated independently and then discussed.

- **norm-referenced testing** Using test results to compare the student with other students from that year and from other years. The result does not give any information about the student’s individual performance, instead it compares him/her to the norm.
- **Oral Proficiency Interview (OPI)** Popular format for the assessment of speaking skills in which the candidate engages in a directed interaction with an interlocutor; developed for the US Government in the 1950s, and influential in many contexts since.
- **proficiency tests** Tests which aim to establish a candidate’s readiness for a particular communicative role, e.g. in a work or educational setting.
- **profiles** These break down marking into a number of different areas so that both teacher and student can see where they are successful and where the reader/listener has difficulties in understanding. Compare **bands**. **Profiles** have several descriptions of abilities for each sub-skill. The technical term for each of these is a **descriptor**.
- **prompt** In the assessment of speaking or writing, the stimulus to which the candidate responds in speaking or writing.
- **raters** Those who judge performance in productive tests of speaking and writing, using an agreed **rating procedure** and criteria in so doing.
- **rating procedure** Agreed procedure followed by **raters** in judging the quality of performances, particularly in the assessment of speaking and writing.
- **rating scale** An ordered set of descriptions of typical performances in terms of their quality, used by **raters** in **rating procedures**.
- **reliability** Consistency of measurement of individuals by a test, usually expressed in a reliability coefficient.
- **rubric** The instructions to candidates before each question or task.
- **self-assessment** The process by which learners are trained to evaluate their own performances, e.g. imagining how well they would cope in a range of real-life settings.
- **stakeholders** Those who have an interest in the development and operation of a test; e.g. funding authorities, educators, test developers, candidates, receiving institutions, etc.
- **summative evaluation** This is evaluation done and feedback gained at the end of (a stage of) a process. It looks at general feedback to the teaching

procedure used, so that next year's course can be changed according to what has been more or less successful.

- **test method** The way in which the candidate is asked to engage with the materials and tasks in the test, and how their response will be scored. (called **test technique** by some authors – see [3])
- **test-taker feedback** The opinions of test-takers, usually at the stage of trialling, on the content and format of test materials; used to revise the

materials.

- **test validation** The process of investigating the quality of test-based inferences, often in order to improve this basis and hence the quality of the test.
- **The Test of English as a Foreign language (TOEFL)** A test of English as a foreign language used to assist in the selection of international students wishing to study at American universities, developed and administered by the **Educational Testing Service**.

BIBLIOGRAPHY

1. BAXTER, A. - *Evaluating your Students*, London, Richmond Publishing, 1997, pp. 94-95
2. MCNAMARA, T. - *Language Testing*, Oxford, Oxford University Press, 2000, pp. 131-138
3. WEST, R. - *Assessment in Language Learning*, Manchester University, 1995