

LECTURE 7. ASSESSMENT AND TESTING

In the context of teaching in general, feedback is information that is given to the learner about his or her performance of a learning task, usually with the objective of improving this performance. Some examples in language teaching: the words 'Yes, right!', said to a learner who has answered a question; a grade of 70% on an exam; a raised eyebrow in response to a mistake in grammar; comments written in the margin of an essay.

Feedback has two main distinguishable components: assessment and correction. In assessment, the learner is simply informed how well or badly he or she has performed. A percentage grade on an exam would be one example; or the response 'No' to an attempted answer to a question in class; or a comment such as 'Fair' at the end of a written assignment. In correction, some specific information is provided on aspects of the learner's performance: through explanation, or provision of better or other alternatives, or through elicitation of these from the learner. Note that in principle correction can and should include information on what the learner did right, as well as wrong, and why! - but teachers and learners generally understand the term as referring to the correction of mistakes, so that is (usually) how it is used here.

The relationship between assessment and correction

It is, of course, perfectly possible to give assessment without correcting, as when a final percentage mark on an exam is made known to a learner without the exam itself being returned or commented on. The other way round is very much less feasible: it is virtually impossible to comment on what is right or wrong in what a learner has done without conveying some kind of assessment. If a correction is supplied, the learner is very aware that this means the teacher thinks something was wrong; if comment is given on why something was appropriate, there is necessarily an underlying message of commendation.

Teachers are sometimes urged to be 'non-judgemental' when giving feedback. Although any meaningful feedback is going to involve some kind of judgement. It is more useful, perhaps, to accept that there is judgement involved, but to try to make the attitude to this more positive: that mistakes are a natural and useful part of language learning; that when the teacher gives feedback on them, the purpose is to help and promote learning; and that 'getting it wrong' is not 'bad', but rather a way into 'getting it right'.

Approaches to the giving of feedback

Below you will find expressions of selected opinions on the nature and functions of assessment and mistake correction; these are based on different theories of language learning or methodologies

Assessment: different opinions

Audio-lingualism

Negative assessment is to be avoided as far as possible since it functions as 'punishment' and may inhibit or discourage learning. Positive assessment provides

reinforcement of correct responses, and promotes learning.

Humanistic methodologies

A crucial function of the giving of assessment is to preserve and promote a positive self-image of the learner as a person and language learner. Assessment therefore should be positive or non-judgemental.

Skill theory

For successful acquisition of a skill, the learner needs feedback on how well he or she is doing; hence the importance of the provision of constant and honest assessment

The correction of mistakes: different opinions

Audio-lingualism

Learner mistakes are, in principle, avoided by the limiting of progress to very small, controlled steps: hence there should be little need for correction. The latter is, in any case, not useful for learning; people learn by getting things right in the first place and having their performance reinforced.

Cognitive code-learning

Mistakes are regrettable, but an unavoidable part of learning; they should be corrected whenever they occur to prevent them occurring again.

Intel-language

Mistakes are not regrettable, but an integral and important part of language learning; correcting them is a way of bringing the learner's 'intwerguage' closer to the target language.

Communicative approach

Not all mistakes need to be corrected; the main aim of language learning is to receive and convey meaningful messages, and correction should be focused on mistakes that interfere with this aim, not on inaccuracies of usage.

Monitor theory

Correction does not contribute to real acquisition of the language, but only to the learner's conscious 'monitoring' of speech or writing. Hence the main activity of the teacher should be to provide comprehensible input from which the learner can acquire language, not to correct.

Assessment

Most of the feedback we give our learners is ongoing correction and assessment directed at specific bits of learner-produced language with the aim of bringing about improvement; the type of assessment involved here is sometimes called 'formative', since its main purpose is to 'form': to enhance, not conclude, a process. Distinct from this is the assessment usually termed 'summative', when the teacher evaluates an overall aspect of the learner's knowledge in order to summarize the situation: how proficient he or she is at a certain point in time, for example, or how much he or she has progressed during a particular course. Summative assessment may contribute little or nothing to the ongoing

Teaching/learning process; but it is a part of the teacher's job, something we need to know how to do effectively.

Below are descriptions of various ways of gathering the information which will serve as a basis for assessment, and of some common criteria used for assessing it.

Gathering information (1): Tests

The most common way of gathering information for assessment is through tests; the usual criterion is an arbitrary level which the learner is expected to have reached; and the result is generally expressed through percentages.

Gathering information (2): Other sources

There are, however, various problems with tests as a basis for summative assessment: they are a one-off event which may not necessarily give a fair sample of the learner's overall proficiency; they are not always valid (actually testing what they say they are) or reliable (giving consistent results); and if they are seen as the sole basis for a crucial assessment in the learner's career, they can be extremely stressful.

Other options do, however, exist. These are summarized below.

1. *Teacher's assessment.* The teacher gives a subjective estimate of the learner's overall performance.

2. *Continuous assessment.* The final grade is some kind of combination of the grades the learner received for various assignments during the course.

3. *Self-assessment.* The learners themselves evaluate their own performance, using clear criteria and weighting systems agreed on beforehand.

4. *Portfolio.* The learner gathers a collection of assignments and projects done over a long period into a file; and this portfolio provides the basis for assessment.

Criteria

Having collected the 'evidence' of the learners' proficiency in one or more of the ways described above, the teacher has to decide how good it is? The following are some of the possibilities.

1. *Criterion-referenced:* how well the learner is performing relative to a fixed criterion, where this is based on an estimation of what it is reasonable or desirable to demand from learners at the relevant point in their development (age, career, level, stage of a course).

2. *Norm-referenced:* how well the learner is performing relative to the group. In this case, a group of slow learners would be assessed according to different, easier, norms than a group of faster ones.

3. *Individual-referenced:* how well the learner is performing relative to his or her own previous performance, or relative to an estimate of his or her individual ability.

What criteria do/would you yourself use in assessing learners' performance? Would you combine different criteria? Would you take into account learners' effort, motivation and progress in deciding on a final grade?

Correcting mistakes in oral work

There are some situations where we might prefer not to correct a learner's mistake: in fluency work, for example, when the learner is in mid-speech, and to correct would disturb and discourage more than help. But there are other situations when correction is likely to be helpful.

The recommendation not to correct a learner during fluent speech is in principle a valid one, but perhaps an over-simplification. There can be places where to refrain from providing an acceptable form where the speaker is obviously uneasy or 'floundering' can actually be demoralizing, and gentle, supportive intervention can help. Conversely, even where the emphasis is on getting the language right, we may not always correct: in a grammar exercise, for example, if the learner has contributed an interesting or personal piece of information that does not happen to use the target form; also, when they have got most of an item right we may prefer not to draw attention to a relatively trivial mistake.

Oral corrections are usually provided directly by the teacher; but they may also be elicited from the learner who made the mistake in the first place, or by another member of the class. Corrections may or may not include a clarification of why the mistake was made, and may or may not require re-production of the acceptable form by the learner.

As important as what the correction consists of is how it is expressed: gently or assertively, supportively or as a condemnation, tactfully or rudely. On the whole, of course, we should go for encouraging, tactful correction; but it is less easy to generalize about gently/assertively: some learner populations respond better to the one, some to the other. In general, in fact, learner responses to different expressions of feedback are often surprising: a teacher correction that seems to an observer a humiliating 'put-down' may not be perceived as such by the learner to whom it was addressed; or an apparently gentle, tactful one may give offence. A good deal of teacher sensitivity is needed here.

5.2 Tests and Testing

People vary very widely in their reactions to tests. Some like the sense of challenge; others find it unpleasant. Some perform at their best under test conditions, others perform badly.

Thus, it would be a mistake to come out with sweeping statements like: 'People get very stressed when they are tested', or 'Tests are unpopular'. The amount of unpleasant stress associated with a test depends on various factors, at least some of which may be under the control of the teacher: how well the learners are prepared for it and how confident they feel of success; what rewards and penalties are associated with success or failure (how important the results are perceived to be); how clear the test items are; how easy the test is as a whole; how often such tests are given; and so on.

Types of tests

1. Questions and answers

These can be used to test almost anything. The more 'closed' the question is (that is, the fewer the possible options for correct answers), the easier the item will be to mark. It is fairly easy to compose and grade closed-ended questions; more open, thought-provoking ones are more difficult, but may actually test better.

2. True/false

This does not directly test writing or speaking abilities: only listening or reading. It may be used to test aspects of language such as vocabulary, grammar, content of a reading or listening passage. It is fairly easy to design; it is also easy to administer, whether orally or in writing, and to mark.

3. Multiple-choice

This may be used for the same testing purposes as true/false items; it does test rather more thoroughly since it offers more optional answers and is obviously very easy to mark. It is administered more conveniently through writing; but note that since the reading of the question-and-options is fairly time-consuming, the process of comprehension of the actual question items may take more time and effort than the point ostensibly tested, which raises problems of validity. Another important problem is that good multiple-choice questions are surprisingly difficult to design: they often come out ambiguous, or with no clear right answer, or with their solutions over-obvious. They are to be approached with caution!

4. Gap-filling and completion

This usually tests grammar or vocabulary, as in the examples. It is tedious to compose, though not so difficult as multiple-choice; it is more easily administered in writing than in speech; the marking is usually simple. You may need to be aware that there is more than one possible right answer.

5. Matching

This usually tests vocabulary, and is rather awkward to administer orally: thus it is best presented written on the board or on paper, though responses may be either oral or in writing. Items can be time-consuming and difficult to compose, and again, there may be alternative 'right' answers to any particular item. Answers are fairly easily checked.

6. Dictation

This mainly tests spelling, perhaps punctuation, and, perhaps surprisingly on the face of it, listening comprehension: people can only usually write words down accurately from dictation if they understand them. It does not, however, test other writing skills or speech, and involves very little reading. It may supply some information on testees' passive knowledge of pronunciation, grammar and vocabulary. It is very easy to prepare and administer; it is relatively easy to mark, though there may be a problem deciding how much weight to attribute to different mistakes.

7. Cloze

This tests (intensive) reading, spelling, and to some extent knowledge of vocabulary and grammar. It can be adapted to 'target' specific language items, by, for example, omitting all the verbs (in which case it is not, strictly speaking, 'cloze', but rather 'gap-filling'). It is fairly easy to prepare and administer. Marking can be tricky: you may find it difficult sometimes to decide if a specific item is 'acceptable' or not.

8. Transformation

This item is relatively easy to design, administer and mark, but its validity may be suspect. It tests the ability of the testee to transform grammatical structures, which is not the same as testing grammar: a testee may perform well on transformation items

without knowing the meaning of the target structure or how to use it in context. Marking is fairly straightforward.

9. Rewriting

This tests the same sort of thing as transformation, but is likely to reflect more thorough knowledge of the target items, since it involves paraphrasing the entire meaning of a sentence rather than transforming a particular item. It is, however, more difficult to compose, and the marking may be more subjective. It is, as its name suggests, usually done in writing.

10. Translation

A technique which, at the time of writing, is for various reasons rather unpopular, but in my opinion undeservedly so. In a monolingual class whose teacher also speaks the learners' mother tongue, the translation of a 'bit' of language to or from the target language can give very quick and reliable information on what the testee does or does not know, particularly when it involves entire units of meaning (phrases, sentences) within a known context. Translation items are also relatively easy to compose - even improvise, in an informal test - and administer, in either speech or writing. Marking may sometimes be more difficult, but not prohibitively so.

11. Essay

This is a good test of general writing abilities. It is relatively easy to provide a topic and tell the class to write an essay about it but marking is extremely difficult and time-consuming. It must be clear in advance, both to you and to the students, how much emphasis you are going to lay on language forms, such as spelling, grammar, punctuation, and how much on aspects of content, such as interest and originality of ideas, effectiveness of expression, organization

12. Monologue

This tests oral fluency in 'long turns' - something not everyone can do in their mother tongue! It also tests overall knowledge of pronunciation, grammar and vocabulary. To choose a topic and allot it is not so difficult; to assess is very difficult indeed, demanding concentration and a very clear set of criteria and weighting system.

Stages in testing. Below are given some recommendations of an experienced teacher how to organize testing.

Before the test

I use the period leading up to the test in order to do all I can to ensure that my students will succeed in it. Thus the tests are announced at least a week in advance in order to give them plenty of time to prepare and details are given of when, where and how long the test will be. The class is also told as precisely as possible what material is to be tested, what sort of items will be used, and how answers will be assessed. I sometimes give them 'test-tips' - for example, how best to allot time, or what to do first - particularly if they are coming near to the state school-leaving exam, for which my course is to some extent a preparation. I usually allow at least some class time for revision, in order to encourage and help with pre-test learning.

Giving the test

It is quite important for me to administer the test myself, and more pleasant for my students. Thus, I will be able, if I wish, to remind them about the test content, format and marking system before giving out the papers; and sometimes run through the instructions with them after doing so in order to make sure that everything is clear - as well as wishing them good luck!

During the test, I may help students who still have difficulty with instructions; I do not normally help with the content itself.

After the test

The tests are marked and returned as quickly as possible (within a week) so that we can discuss specific points while the test is still fresh in the students' minds. Usually I will go through the answers in class, but fairly briskly; points that seem to produce special problems I note for more leisurely re-presentation and further practice in the future. I do not usually ask students to copy out corrected answers: this is, I think, more tedious than helpful for them. It is better and more interesting to provide the practice in the same language points in other activities, using new content and tasks.

Types of assessment - some definitions

Formative Assessment

This occurs in the short term, as learners are in the process of making meaning of new content and of integrating it into what they already know. Feedback to the learner is immediate (or nearly so), to enable the learner to change his/her behavior and understandings right away. Formative Assessment also enables the teacher to "turn on a dime" and rethink instructional strategies, activities, and content based on student understanding and performance. His/her role here is comparable to that of a coach. Formative Assessment can be as informal as observing the learner's work or as formal as a written test. Formative Assessment is the most powerful type of assessment for improving student understanding and performance.

Examples: a very interactive class discussion; a warm-up, closure, or exit slip; a on-the-spot performance; a quiz.

Interim Assessment

This takes place occasionally throughout a larger time period. Feedback to the learner is still quick, but may not be immediate. Interim Assessments tend to be more formal, using tools such as projects, written assignments, and tests. The learner should be given the opportunity to re-demonstrate his/her understanding once the feedback has been digested and acted upon. Interim Assessments can help teachers identify gaps in student understanding and instruction, and ideally teachers address these before moving on or by weaving remedies into upcoming instruction and activities.

Examples: Chapter test; extended essay; a project scored with a rubric.

Summative Assessment

This takes place at the end of a large chunk of learning, with the results being primarily for the teacher's or school's use. Results may take time to be returned to the student/parent, feedback to the student is usually very limited, and the student usually has no opportunity to be reassessed. Thus, Summative Assessment tends to have the least impact on improving an individual student's understanding or performance. Students/parents can use the results of Summative Assessments to see where the student's performance lies compared to either a standard (MEAP/MME) or to a group of students (usually a grade-level group, such as all 6th graders nationally, such as Iowa Tests or ACT). Teachers/schools can use these assessments to identify strengths and weaknesses of curriculum and instruction, with improvements affecting the next year's/term's students.

Formative assessment

Formative assessment is an integral part of teaching and learning. It does not contribute to the final mark given for the module; instead it contributes to learning through providing feedback. It should indicate what is good about a piece of work and why this is good; it should also indicate what is not so good and how the work could be improved. Effective formative feedback will affect what the student and the teacher does next.

Summative assessment

Summative assessment demonstrates the extent of a learner's success in meeting the assessment criteria used to gauge the intended learning outcomes of a module or programme, and which contributes to the final mark given for the module. It is normally, though not always, used at the end of a unit of teaching. Summative assessment is used to quantify achievement, to reward achievement, to provide data for selection (to the next stage in education or to employment). For all these reasons the validity and reliability of summative assessment are of the greatest importance. Summative assessment can provide information that has formative/diagnostic value.

'Authentic' or work-integrated assessment

'Authentic' or work-integrated assessment is an assessment where the tasks and conditions are more closely aligned to what you would experience within employment. This form of assessment is designed to develop students skills and competencies alongside academic development. The [Collaborate project](#) at Exeter developed a set of tools to support academic staff in the design of authentic assessments, including a [dimensions model](#), [iTest](#) and associated [Tech Trumps](#). There is also an online Assessment Designer available which will allow you to design an assessment using a PC or tablet device.

[Launch the Assessment Designer†](#)

Diagnostic assessment

Like formative assessment, diagnostic assessment is intended to improve the learner's experience and their level of achievement. However, diagnostic assessment looks backwards rather than forwards. It assesses what the learner already knows and/or the nature of difficulties that the learner might have, which, if undiagnosed, might limit their engagement in new learning. It is often used before teaching or when a problem arises.

Dynamic assessment

Dynamic assessment measures what the student achieves when given some teaching in an unfamiliar topic or field. An example might be assessment of how much Swedish is learnt in a short block of teaching to students who have no prior knowledge of the language. It can be useful to assess potential for specific learning in the absence of relevant prior attainment, or to assess general learning potential for students who have a particularly disadvantaged background. It is often used in advance of the main body of teaching.

Synoptic assessment

Synoptic assessment encourages students to combine elements of their learning from different parts of a programme and to show their accumulated knowledge and understanding of a topic or subject area. A synoptic assessment normally enables students to show their ability to integrate and apply their skills, knowledge and understanding with breadth and depth in the subject. It can help to test a student's capability of applying the knowledge and understanding gained in one part of a programme to increase their understanding in other parts of the programme, or across the programme as a whole [1]. Synoptic assessment can be part of other forms of assessment.

Criterion referenced assessment

Each student's achievement is judged against specific criteria. In principle no account is taken of how other students have performed. In practice, normative thinking can affect judgements of whether or not a specific criterion has been met. Reliability and validity should be assured through processes such as moderation, trial marking, and the collation of exemplars.

Ipsative assessment

This is assessment against the student's own previous standards. It can measure how well a particular task has been undertaken against the student's average attainment, against their best work, or against their most recent piece of work. Ipsative assessment tends to correlate with effort, to promote effort-based attributions of success, and to enhance motivation to learn.

