3. Evaluation and In-company Language Training

Peter Strutt, The British Institute

Abstract
Unless clear evaluation procedures are laid down in advance, the sponsors of in-company language training and the trainees themselves may end up disillusioned with the investment made in language learning. There may be the feeling that a lot of time and money has been spent for little return, at least in a tangible, measurable form. It is therefore essential to ensure that clear objectives are defined at the outset and that the achievement of these objectives is monitored at all stages. This implies good testing procedures for: (1) testing newly recruited executives to ascertain language level; (2) initial placement testing, when needs are assessed, trainees are chosen for level (in relation to the occupational language required) and diagnoses made on the content goals of the training; (3) progress (formative) testing to ensure that input becomes intake and, if necessary, to take corrective action; (4) achievement (summative) testing to measure the effectiveness of the course programme as a whole and to point the learners in future directions; (5) periodic reassessment of language during an individual's career, for example when an assignment requires a specific level in a language, or for a change in responsibilities. This paper will demonstrate how the "washback" effect of evaluation can lead to changes in the design and organisation of in-company programmes.

The return on investment

C ompanies and other organisations which pay for language training need to ensure that the results obtained are 'value for money'; in other words, the return on investment should be visible and measurable. For a training manager, the resources devoted to any kind of training should be cost-effective; if they are not, the financial department may wish to cut the budget!

The benefits that accrue from an 'intangible' service such as language training are notoriously difficult to quantify. Nevertheless, I believe it is possible to take a number of measures which will go some way to ensuring that in-company language training stands a good chance of success and that such success can be demonstrated.

Language training at the place of work should be a partnership between three sets of people: the training and human resources personnel, the language services

provider and the language trainers, and the trainees themselves.

The training and human resources personnel

If language training is to be fully effective, training and human resources personnel need to draw up an action plan at three levels: needs analysis and evaluation procedures, staffing and resources, and the integration of language training within a global training policy.

Needs analysis and evaluation procedures

A fundamental consideration is the selection of potential trainees. Language training will be more effective if there is a real need (present or future) and opportunity for the skills acquired to be put into practice immediately. As a result, the decision to send an individual on a course must be taken according to well-defined criteria, preferably involving the trainee him/herself, the trainee's immediate hierarchical superior and the service provider.

This needs analysis should also be based upon a wider analysis of the language-related needs of a particular company division, a factory, a specific sector of activity, a job description or a project. Such an analysis should enable an organisation to better target the kind of outcome the language trainer should work towards and provide the language school and trainer with a precise set of specifications, together with targeted goals.

In my experience this rarely happens, yet an initial evaluation by the human resources department of its real language needs would seem to be an important pre-condition for success.

Staffing and resources

For a course to work properly, staff being trained need to be released, need to be present in class and not called out urgently or forced to cancel. Companies should provide at least some of the necessary tools and equipment, for example a self-access centre.

The integration of language training within a global training policy

Language training stands a better chance of success if it is considered as being on a par with other kinds of job training and not viewed as a poor relation or a form

of relaxation. Where possible, language training should conform to any company policy on training and be subject to the same or similar control procedures.

The language services provider and the language trainers

Both the director of studies and the trainers should be involved in setting out training goals and establishing evaluation procedures at the outset of a course. This should involve consultation with the organisation and the potential trainees and set out a programme of work whose success can be monitored and measured. If the goal of a 60-hour course is to improve telephone communication skills, presentation techniques, and negotiation strategies, then there needs to be a way of determining at the outset what the present language skills of the course participants are and a way of measuring improvement (if any) as a result of the course.

The training methods, course books etc. can then be selected as the best means to attain those course goals for those particular course participants. This is in itself a form of evaluation as all teachers are intimately involved in assessing the suitability of materials for their own classes.

The trainees themselves

The trainees need to be briefed on the content of the training, the methodology, the purpose, and the evaluation. They need to be aware of the amount of effort involved in reaching the stated goals and their own role in their attainment.

A suggested procedure for ELT

The following is a possible procedure for implementing an in-company course and evaluating both the course and the trainees.

1. **Needs are analysed** by the training manager or immediate superior in the hierarchy, in consultation with the individual who has requested language training. These needs may be of an individual nature or common to a professional profile (language for technicians, language for receptionists, language for financial controllers etc.). In the latter case, a list of activities linked to the job in question will help to clarify the needs.

2. **Target performance levels are established** for each of the language skills (Listening, Speaking, Reading, Writing) with a set of prose descriptors, and the means and timescales involved in attaining these objectives.

3. **A profile of the trainee's initial language level is drawn up** using a placement test. A comparison is made between his/her present competence as measured and the level to be attained. If there is a difference between current level and target level as defined by needs analysis for occupational performance, then there is a **training gap** which needs to be filled.

If, for example, an individual needed to be able to use the telephone on a regular basis in English and had taken a placement test which revealed that he /she was at this level (say level 2)

S/he can understand simple messages or instructions when given slowly and clearly.

S/he can respond to predictable questions and express basic factual information (e.g. prices, times, dates, places, destinations, names, quantities, etc.).

S/he can pronounce most individual sounds acceptably but stress and intonation patterns are unnatural. Fairly frequent errors in usage may irritate the interlocutor and Ll speech features frequently impair communication.

More complex discourse at normal speed requires excessively frequent repetition and reformulation from the interlocutor.

then logically the training programme would concentrate on extending this person's ability by focusing on an improvement of stress and intonation and lexical and grammatical accuracy as they affect communication, using phone conversations and the like which go beyond the mere expression of basic information.

After the course has finished a similar test could be administered and hopefully the person would have better listening skills, be demonstrably better at coping with more complex messages, and speak using more acceptable stress and intonation patterns according to a definition of a superior level (level 3).

Level 3

S/he can easily understand simple messages or instructions and can cope with more complex discourse and unpredictable situations at normal speed without requiring excessively frequent repetition and reformulation from the interlocutor.

S/he can pronounce most individual sounds correctly, and stress and intonation patterns are by and large natural. Errors in usage may still occur but do not unduly irritate the interlocutor or impede communication

What is important is to make sure that, insofar as possible, the test used actually provides evidence that can be matched against the language levels described in the language skill framework. This is by no means an easy feat but it is our job to make sure that the test does demonstrate what we want it to demonstrate.

4. **A training programme can then be drawn up** taking into account current and target levels, needs, individual learning preferences, timescales, finance, motivation and so on. The design of the course will be realised with relation to the previously agreed training goals. Depending on circumstances, this programme may be in a group, one-to-one, or even on a self-access basis.

In most cases, the training should draw upon a common core of language functions, grammar, lexis and skills work linked to topics relevant to the work situation.

5. **The training programme is implemented and modified** as and when necessary. During the programme the individual's progress is measured by using (or creating) a suitable progress test, and any remedial work is scripted into the course to ensure the achievement of the objectives previously defined. The trainees also have the opportunity to influence the direction of the course and the trainer should also be taking stock.

6. At the end of the programme **a final achievement test is administered** which, together with the appraisal of the tutors/trainers concerned, establishes the extent to which the course has been effective.

Ideally, an appraisal of the effectiveness of the training should be carried out in the work environment to ensure that skills rehearsed during training have been transferred to the professional context. This could take the form of a self-appraisal.

7. Whatever the outcome, there will always be a need for **maintenance** of the language and skills which have been acquired. Language ability, like other skills, with time tends to fade away unless there is opportunity for continuing use of previously-acquired skills. The maintenance of language skills should thus be a major priority for the individual and the company.

So, according to this procedure, evaluation drives the whole process:
*   needs are evaluated
*   the target language performance to be reached after training is specified
*   the potential trainee's language level is assessed
*   the degree of attainment of the target performance is measured both during and after completion of the course
*   in a process known as "washback", teaching programmes are positively influenced by effective evaluation.

Testing instruments: design features

However, this procedure does depend on adequate testing instruments, since the attribution to any individual of a particular level presupposes tests which adequately measure the job-related skills that were defined at the outset.

Therefore the methods used to assess competence should be:

valid
i.e. individuals should have an opportunity to demonstrate the skill being assessed (it would be unreasonable to expect a recital of the Highway Code as evidence of an ability to drive a car)

reliable
produce similar results with similar individuals on separate occasions and/or with different assessors. This is easy with objectively scored tests but less so with subjectively assessed tasks.

criterion-referenced
they should judge an individual's ability against an established standard and not one person's performance against another's (norm-referenced)

qualitative
provide an interpretable result in terms of profile reporting rather than the spurious objectivity of a numerical score

manageable
not make excessive demands in terms of administration or time

task-oriented
the focus of assessment should always be on a behavioural **outcome** — what has actually been achieved (or not) through the individual's use of language — rather than a focus on the language itself.

ethical
testees should be informed that they are being tested, and told what they are being tested for and what will be done with the results.

Above all, no test-taker should be harmed by the test.

When working in-company I have found that it is advisable to devise tests with various sections, each section testing a particular skill in a **direct** way with the tasks set as open-ended as possible, thereby enabling test-takers to perform at their own level.

It is then possible to use a scale of performances for each skill (for example: a Professional Writing Scale, a Professional Speaking Scale). If you test all the skills you might get an individual who is good at speaking, less good at listening, reasonably good at extracting meaning from a written text but hopeless at writing a document in a professional context. This individual might be coded (on a scale of 6) as

<div align="center">

S4   L2   R3   W1

</div>

This does not necessarily imply a training gap for writing because the individual in question may never have to do any. But it does suggest that s/he should concentrate more on listening skills instead of speaking too much!

Direct tests would do the following
reading skills
focus on different types of reading — reading for general meaning, reading for detail, recognising cohesive links in texts.

listening skills
focus on things like phone conversations, the recognition of numbers, addresses, message-taking and the understanding of longer stretches of discourse with an 'argument' to follow and understand.

writing skills
measure the test-taker's ability to respond to requests for information sent by fax, or write a reply to a complaint, using prompts.

speaking skills
measure the individual's ability to communicate basic information, present a case, describe a product, react to another point of view etc.

A checklist
Not everybody is involved in test design (though we all should be if we say we are meeting companies' and individuals' specific needs and customising our training to achieve that). However, most people are involved in test implementation, so here is a series of maxims which I have found useful to bear in mind.

1. There is no point in giving a test if you do not need or will not use the information, if you cannot interpret the results, or if you will not believe them if they go against your expectations.
2. Achievement tests should relate to what has been taught / learned during training, and also to the way it has been taught. For example, it is probably not reasonable to expect trainees to perform spontaneously with fluency in oral role-plays if the course has primarily focused on listening skills.

3. Tests will always have an effect on the teaching and the syllabus (the **washback effect**). Try to ensure that a test will have a beneficial effect by testing important, meaningful skills / abilities / knowledge.

4. Tests should be **valid**: they should test what you really want to measure and find out. The test designer should be clear about what exactly s/he wants to test, and as far as possible ensure that the test measures that and not irrelevant abilities or knowledge. For example, minor problems of spelling of responses should not be taken into account in a test of listening comprehension.

5. Tests should be within the reach of the intended population — those trainees who know the language or have the ability being tested should perform well on the test.

6. The targeted items should not demand general knowledge or skills that are not directly related to language: for example, irrelevant manage-ment skills such as the ability to conduct an interview, etc.

7. Tests should be practicable — not require too much equipment or special conditions, cost a lot of money or take up too much time. Testing should not take up a disproportionate amount of time, at the expense of teaching.

8. Tests should be as **reliable** as is practical: we need to be sure that the results are believable — that similar results would have been obtained on similar occasions by the same trainees, even with different markers.

9. Always pilot a test: other people will see ambiguities where the test designer only saw one possible interpretation. Always try out a test on colleagues, and if possible on a control group, to see if they can understand how to do the test, and if they produce unexpected responses or interpretations.

10. Test instructions (rubrics) should be clear, precise and simple. If they have to be complex or long, it is better to give them in the native language. Where possible, give a clear example before the test proper. Testees should not only know what they are expected to do and how, but should also have an idea of how they will be judged i.e. what the marker will be looking for.

11. Define the criteria for marking, where possible in advance of the testing. Try to ensure agreement on criteria among markers.

12. If unexpected responses are given, different markers should judge these in similar ways and should be sufficiently flexible to allow the correct but unexpected response, if it is appropriate in a professional context.

13. Correctors should agree on how performance is to be judged: are they looking for grammatical correctness? Or the appropriate use of lexis? Or good arguments? Or logical organisation of ideas? Or complex sentences? Or good handwriting and layout?

14. Since language is used in context for a purpose, avoid decontextualised test items. Provide sufficient context for adequate disambiguation.

15. The context should clearly require the answer the tester has in mind, and others should be implausible.

16. The tester needs to foresee all possible answers, especially in open-ended items.

17. Remember that multiple-choice questions are not necessarily easier to answer than open-ended questions. They are certainly more difficult to design well.

18. In multiple-choice questions, the distracters should distract, and for relevant reasons. For example, the options should be of equal length, or paired by length, and of similar difficulty, or paired by difficulty.

19. Avoid irrelevant bias like test-wiseness : knowing how to apply a particular test technique (e.g. multiple choice).

20. Avoid bias against people with different cultural behaviours, beliefs or knowledge.

21. Avoid introducing extraneous skills such as good handwriting or mathematical ability, which are unrelated to language proficiency.

22. Test items should not be biased in favour of individuals with particular subject or background knowledge, unless this is what is being measured.

23. Write test items which will collect an adequate sample for measurement purposes.

24. Encourage the creative use of the language in your test items.

25. Test tasks should be as similar as possible to the real world communicative use of language. For example, a communicative task like writing should give the trainee something to write to someone (not the examiner) for a valid reason.

26. Beware of marking as incorrect any responses which are possible in some regional variety of English, or possible in the spoken form and not the written form.
27. The language of rubrics or comprehension questions should be easier than the language of the texts.
28. It should not be possible to answer reading or listening items correctly without the written or spoken text.
29. Testing should be seen in a humanistic light with tests written and administered so as to minimise anxiety on the part of the testees.
30. Tests should be presented in a positive light, not as sanctions.

Number 30 raises, however, an interesting ethical question. Suppose two people with more or less the same profile are competing for an important posting in a subsidiary abroad where English is the language of day-to-day communication.

The company approaches you, the business English teacher, and asks you for your professional judgement on their respective competences. Here you are suddenly involved in making important decisions which could affect the whole of a person's subsequent career.

Conclusion

This is why it is important for us all to be aware of the issues involved in evaluation and to make sure that we perform this professional duty effectively. If we direct people towards inappropriate tests or make ill-informed judgements on the basis of our own poor test design then we are not doing our jobs properly.

Peter Strutt has been in France since 1979. At the British Institute in Paris he is director of the distance learning courses and does teacher training in TEFL foundation courses and the UCLES Delta programme. He teaches as well on an MA in Oral Testing course and is an assessor for Cambridge examinations. He has worked on a regular basis in the engineering and management schools of the Institut National des Télécommunications. His interests are evaluation, self-directed learning, professional English, materials writing and translation. His publications include *"Longman Better English Usage"*, *"English for International Tourism"* (Longman) *"Powerhouse workbook"* (Longman). *"Words at Work"* (CUP), *"English for Translation"* (Belin), *"English for Advertising & Communication"* (Belin).