This article advocates making the provision of opportunities for learning the main objective of language testing. It recognizes the need for tests to be fair, valid, and reliable, but asserts the priority of what it calls ‘learning validity’, in order to prevent time being wasted on language courses on tests, and the preparation for them. The article lists what can be learnt from tests with learning validity, and then discusses why such benefits are rare. It surveys the literature on testing in an attempt to find support, and then provides examples of ways of testing which could help language learners to learn from being tested. In particular, it looks at how learners could really benefit from their preparation for tests, from taking tests, and from constructive feedback whilst and after taking them.

Introduction

In my view the main purpose of language testing is to provide opportunities for learning, both for the students who are being tested, and for the professionals who are administering the tests.

The students can learn from:

- the work they do with the teacher, and by themselves in preparation for the tests
- the opportunities which arise during the tests for developing what they know and what they can do
- the feedback which they receive after the test, both from their own reflection and from professionals who have monitored their performance on the tests

The professionals can learn from the performance of the students during their preparation for the tests, from their test performance itself, and from their response to post-test feedback. They can gain information, for example, about:

- what the students know and what they do not know (both explicitly and implicitly) about the target language
- what the students can and cannot do in the target language
- what progress the students are making
- what effect the teaching has had on the students
- what the students are likely to be able to do with the language in a target context
- what the students need to learn
- what can be done to help the students to learn
- what the students know and can do in comparison with other students.
While it is obviously important that tests should be fair, valid, and reliable, the most important of all is that tests should provide useful opportunities for learning.

**Formative vs. summative tests**

A common complaint of teachers preparing students for end of course summative examinations (whether they be global, national, or institution specific) is that the examinations might be valid and reliable, but the students learn very little of communicative value during their preparation for them and while taking the examinations. In recent years many of the global examinations have become more communicative in what they assess, and their washback effect on the classroom has become more positive. However, it is still true to say that, ‘Language testing occurs in an educational and social setting, and the uses of language tests are determined largely by political needs.’ (Bachman 1990: 291). Major summative examinations must be reliable and fair, and must discriminate between the candidates; therefore their means of assessment are still predominantly objective, and they still rely to a large extent on such task types as multiple choice, transformation, substitution, and matching. For example, Paper 1 Reading of the UCLES Certificate in Advanced English (13 June 2001) asks candidates to match information to sections of a review of CDs, to insert paragraphs into gaps in a magazine article, to answer multiple choice comprehension questions on a magazine article, and to match statements to sections of a newspaper article. It does not ask the candidates to give their views or interpretations, or to connect what they read to what they know and think. It does not ask them to read for any purpose other than to answer questions on a text. And it does not ask them to do anything which we would normally do when not reading for a reading examination. It would be possible to find interesting ways of helping learners to prepare for such examinations, but my point is that the candidates will learn nothing about reading or about the language whilst taking such examinations, and gain nothing from practising such examination tasks other than examination readiness.

The purpose of this article is not to criticize the major summative examinations, and I acknowledge the difficulty of designing assessment tasks which replicate ‘real’ use of language, whilst ensuring reliability and fairness for a huge number of candidates. Unfortunately, though, the major summative examinations are so powerful that they tend to determine the procedures not only of formative tests used on students preparing for them, but also on formative tests in general, as well as the content and format of textbooks, and the focus of much of the teaching. You only need to try to write communicative textbooks for examination level classes to realize the power such examinations have in influencing Ministries of Education and publishers to insist on the inclusion of examination type questions which have little relevance to the real world of communication. I speak from recent experience of writing a secondary school coursebook for a Singapore publisher who understandably insisted on the inclusion of discrete comprehension questions and word specific summaries in every unit to help the students to prepare for the Cambridge ‘O’ level examination they would eventually take.
In a published comparison of the FCE and TOEFL examinations (Bachman et al. 1995) there is no mention at all of the usefulness of the examinations in relation to their potential for the provision of opportunities for learning. If it is true that such public summative examinations provide little opportunity for useful learning, then it follows that the preparation for them will also provide little opportunity for learning, and the feedback on them (if any) will be of little value too. What is needed is for public summative examinations to aim to provide useful opportunities for learning, as well as to be fair, reliable, and valid. Only then will formative tests (i.e. those giving information about a learner’s progress during a course) and exam preparation teaching provide useful opportunities for learning. This is what happened, for example, in 1981 when the Ministry of Education in Vanuatu decided to introduce a radically different Primary School Leaving Examination. Instead of the usual multiple-choice, of filling in the blank and essay writing type items, the new examination would set the pupils communicative tasks. Of course, there were no suitable coursebooks to help the teachers to prepare the pupils, and there were no past papers to practise taking. What happened was that workshops were held for teachers at which communicative methodology was introduced and demonstrated, and at which the teachers themselves wrote materials for a book of communicative tasks (Tomlinson 1981). Almost overnight, the teaching and the testing in primary school English lessons changed, and the pupils started to do more useful things in the classroom. A similar story of how changing an examination had beneficial effects on the syllabus, materials, and teaching is told by Hughes (1989: 1).

Obviously though, it is much easier for a small country like Vanuatu to make dramatic changes to its summative examinations, and we cannot expect the major global examinations to do the same. Until ways are developed of examining huge numbers of candidates from many different countries which are not only reliable and fair but which are also valid and realistic, our main concern should be to convince ministries, publishers, and teachers of the value of formative tasks designed to promote useful learning. And we also need to convince them that using such tasks is not going to disadvantage learners when they take their summative examinations.

There is an extensive literature on testing, which has gone a long way to clarifying the differences between evaluation, assessment, and testing, to establishing the important differences in objectives and procedures between formative tests and summative tests, and to helping teachers and examiners to set tests and examinations which are fair, valid, and reliable. However, most books and articles on testing give little space to discussing ways of ensuring that tests and examinations provide opportunities to learn.

In a survey of some of the literature on testing I could find no article or book which directly addressed the issue of testing to learn, though admittedly there was a lot of useful discussion of the need to ensure positive washback from the testing to the teaching (e.g. Hughes 1989). I did, however, come across the following points, which could be interpreted as relating to the issue of providing opportunities to learn:
‘Oral tests must treat people as human beings … we can make a test challenging, instructive and even an enjoyable experience’ (Underhill 1987: 6).

‘It is sensible to make the test as similar to the class materials as possible’ (Buck 2001: 201).

It is possible to reduce the difference between learning and testing by using samples of typical performances rather than one-off tests as a means of assessment (e.g. Alderson 2001).

Tests should be authentic in that the test-taker should be required to deal with genuine texts in ways which correspond to their use in contexts of communication (e.g. Buck 2001; Brindley 1998; Weir 1990).

Inviting the learner to participate in self-assessment can help the learner to become more aware of what they need to learn, and can provide useful insights for the testers too (e.g. Alderson 2001).

It can be useful to the learners to test their ability to use communication strategies (e.g. Read 2001).

‘A good test is one that allows us to make useful inferences about the test taker’s communicative language ability’ (Buck 2001: 92).

‘If we test directly the skills that we are interested in fostering, then practice for the test represents practice in those skills’ (Hughes 1989).

Positive washback (or the effect of testing on instruction) occurs ‘when the testing procedure reflects the skills and abilities which are taught in the course’ (Bachman 1990: 283) and when ‘the effect of teachers teaching for the exam will lead to classroom activities which are in line with the pedagogic approach of the institution’ (Baker 1989: 105–6).

These are all useful points, but the reason why testing to learn is not given greater prominence in the literature on testing is implied in the following statement from Alderson (2001: 203):

The primary purpose of a teaching/learning task is to promote learning, while the primary purpose of an assessment task is to collect relevant information for purposes of making inferences or decisions about individuals—which is not to say that assessment tasks have no potential for promoting learning, but simply that that is not their primary purpose.

This point is made even more strongly in Hughes (1989: 7) when, despite arguing a powerful case for ensuring positive washback, he makes no mention of providing opportunities for learning in his list of the purposes of testing. Likewise Baker (1989) pleads for positive washback, but does not include ‘Does it provide opportunities for learning?’ in his seven key questions about tests.

My argument is that the most effective way to prevent assessment tasks from inhibiting the learning process is to make the promotion of learning their primary purpose (whilst making sure they achieve their assessment purposes too).

Helping learners to learn

I would like to focus on ways of using tests to help learners to learn because much of the existing literature focuses on ways of helping others to use tests to learn about the learners.
Ideally, for an informal test of what the students can do in the target language, the students should not even know when they are going to be tested, and should receive no formal preparation for the test. The test type should be the same as one of the communicative task types that the students are familiar with, and all communicative tasks done as classwork or homework then become informal preparation for the communicative tests. That way students do not waste time in special preparation for an unfamiliar test task, and they are more likely to provide a typical sample of their use of the language than if they are tense and nervous from test anticipation. This is what I did with students of English at Kobe University in Japan, who never knew which tasks were being used as tests, and which were being used as preparation for tests. Thus, a task inviting each student to give me instructions on how to get to a good seafood restaurant from the university could be used as a practice task or a test. The same could be true of tasks in which students made paper airplanes from my instructions, or drew pictures to represent a story I had told them. Over the semester the students did ten such tasks as tests without knowing it, and the marks gained on the tests then made up 50% of the total grade for the semester. Some people might object that such an approach is unfair, as the students have no opportunity to prepare for the tests. My answer would be that each student has exactly the same opportunity, and that it is much fairer to test typical performance than to treat stressed performance as typical.

While it is obviously not possible to conceal the dates of formal examinations, it is possible to replicate classroom task situations in the examination. In that way we can ensure that class preparation for performance examinations is useful, and that if classroom tasks typically replicate features of real-world communication, the washback effect of the examinations will be positive. In the Kobe University end of year summative examinations, for example, I included such classroom features as pre-reading activities, group preparation, and pair writing, and, in some tasks, I even encouraged students to seek help from peers and from the invigilator (on the grounds that such help-seeking is an important skill in communication). In order to achieve as much fairness and reliability as possible, I marked the examinations using broad bands of marks, each defined by detailed performance criteria.

If the test, either formal or informal, is of the learners’ explicit knowledge of the language (either declarative or procedural) then it is both fair and useful to announce the test in advance, and to spend time helping the students to revise and extend what they know. This preparation is time well spent if the test assesses knowledge in ways that relate to real-life use of language (e.g. using knowledge to help produce a draft during a process writing activity; planning strategies to be used in a problematic meeting; monitoring and improving a written text). Another type of test for which the preparation would be extremely useful would be a test requiring students to make use of reference materials (e.g. making use of a dictionary or thesaurus or grammar book in composing and/or monitoring a text).
The main point is that if a test provides useful experience to the students, then preparation for it will be useful too. On the other hand, if the test assesses knowledge or skills of little relevance to real-world communication (e.g. through questions on the grammar of discrete bits of a text, or on the correct word to fill in a blank in a context free sentence) then the preparation for the test could take up time which would be more usefully spent providing learning opportunities of real-world relevance.

Tests are not normally viewed as events in which new learning takes place. But it is perfectly possible for learners to gain new knowledge, and to develop new awareness and skills whilst actually taking a test. This can be achieved by:

- Presenting the learners with new knowledge during the test and then asking them to apply it (e.g. teaching elementary learners about the interrogative during a test and then asking them to design a questionnaire).
- Teaching the learners a new strategy during the test, and then asking them to apply it (e.g. teaching the learners ways of scanning a text for specific information during a test, and then giving them a short time to find information from a text).
- Teaching new language whilst testing something different (e.g. giving a comprehension test on a text teaching a feature of the language).
- Teaching new skills whilst testing something different (e.g. teaching half the class about the skill of visualization when reading or listening and testing them on their ability to give advice on visualization to the other half of the class in groups of four).
- Testing the learners’ ability to use a skill in a novel context (e.g. testing learners who have given short oral presentations on their hobbies or on their ability to give presentations to a group of potential customers).
- Testing learners through a longitudinal project which requires the learners to learn new language and develop new skills in order to complete their project (e.g. a project which involves the learners in groups designing a new product, convincing a company to invite them to present their product, doing a presentation on their product and designing newspaper and TV advertisements for the product).
- Testing the learners’ ability to develop their awareness of the functions of a language feature which they have not been taught (e.g. testing elementary learners on their ability to develop awareness of some of the functions of the passive voice from their exploration of authentic samples of the passive in use).
- Instead of using topics which all the students are familiar with as a basis for reading and listening tests (the normal means of ensuring fairness), using topics which are new to everybody.
- Testing the learners in the context of an experience which is novel to all of them (e.g. taking the learners to the theatre for the first time, and then testing them on their understanding of the play).
- Testing learners in pairs and groups so that they learn from each other whilst doing the test.
Obviously such tests are easier to design and administer as formative tests in the classroom, and it would be very difficult to adapt them for major global examinations. However, it is possible to devise ways of using them in formal testing situations in which fairness, reliability, and validity, are important (e.g. by constructing clear and specific criteria prior to the test, and by focusing on specific aspects of performance during holistic tasks). This is what I did in the end of year examinations at Kobe University.

The most obvious way in which learners can learn from tests is by receiving focused and constructive feedback during and after a test. Unfortunately, though, learners are often given little more than a grade and a bit of general advice (e.g. ‘Be more careful in your use of tenses’) and often they are given no feedback from summative examinations at all.

One way of helping learners to give themselves useful feedback is to include a monitoring and reflection activity as the final part of a test, as in the following example:

Q8
1 What do you think you have done well in this test? (4 marks)
2 What problems have you had doing this test? (4 marks)
3 What are you going to do to help yourself to overcome some of these problems? (4 marks)
4 What have you learnt whilst doing this test? (4 marks)
5 What are you going to do to help yourself to learn more from this test? (4 marks)
(Total: 20 marks)

Other ways of helping learners to gain useful feedback include:
- Giving learners a list of criteria for assessment before they take a test, and then giving them the assessors’ answers to the questions after the test has been marked.
- Giving each learner very focused teaching and advice in response to just one of the problems identified by the marker of a test (this could even be done as feedback to large summative examinations by using pre-prepared feedback cards focusing on predictable problem areas).
- Giving each learner feedback and advice during a test.
- Giving learners generalized and individual feedback after a dummy test, and then assessing them on a similar test.
- Getting learners to monitor (and maybe even grade) their own or other learners’ answers with the aid of criteria and/or reference material.
- Getting learners to monitor (and maybe even grade) their own or other learners’ answers after experiencing further teaching relevant to the focus of the test.

The ideas above for helping learners to learn from tests are not all easily applicable to major global examinations, but all of them can be used in formative testing, and all of them could be used during preparation for a major summative examination. I am convinced from my own experience
that such preparation is actually beneficial to candidates for global examinations, provided that they are given opportunities to practise doing examination like tasks immediately prior to taking the summative examination. But of course, we need empirical evidence that this is so in order to persuade ministries, publishers, and teachers to focus more on assessment tasks from which learners can learn.

Conclusion

Much time and expertise has quite rightly been devoted to the development of ways of ensuring that tests and examinations are valid in that what is tested is worth testing, that they test what the testers want them to test, and that they conform to the expectations of those involved in the testing process. For example, Bachman (1990) has a very useful chapter on test validity (but makes no mention of usefulness for learning as a feature of validity). But perhaps it would help the learners most if a new form of validity was focused on, if time and expertise were devoted to ensuring that learning validity was achieved (i.e. that the test or examination provided opportunities for learning). Then there would be a good chance that the textbooks and teaching would change too, and that at long last, learners would benefit from testing to learn.

References


The author

Brian Tomlinson is Reader in Language Learning and Teaching at Leeds Metropolitan University. He has taught and tested students in Japan, Indonesia, Nigeria, Singapore, the UK, Vanuatu, and Zambia, and has also worked as a teacher trainer, curriculum developer, and university lecturer. He is the Founder and President of MATSDA (the Materials Development Association) and has published many articles and books, e.g. (with Rod Bolitho) Discover English (Macmillan), Materials Development in Language Teaching (Cambridge University Press), Developing Materials for Language Teaching (Continuum Press); (with Hitomi Masuhara) Materials Development (RELC), and Openings (Penguin).

Email: B.Tomlinson@lmu.ac.uk