

# Research Paper

Theme: Types and writing principles of tests

Hakobyan Lusine

Yerevan  
2021

# 1.Introduction

This research paper deals with the types and writing principles of tests. Whether you like them or not, tests are a way of checking your knowledge or comprehension. They are the main instrument used to evaluate your learning by most educational institutions. Most teachers contend that testing is all too often viewed as a necessary evil. One should also note that testing on a frequent and regular basis provides the teacher with valuable information concerning areas of difficulties the learner encounters. In this way the teacher gets more about what aspects of language need further clarification and explanation and subsequently devise remedial activities. It also gives an objective evaluation of a learner's progress individually. In sum , one should view testing as a bridge –building process between teaching and learning and classroom tests as mirrors in which teachers and students see their reflection clearly. ( Valette 1977 )

Testing has evolved and has become extremely prevalent in our educational system today. Language tests are broadly classified into two types: testing skill and testing knowledge of content. Skills such as listening, speaking, reading and writing and sub- skills such as comprehension, vocabulary, grammar ,spelling, punctuation etc. In educational terms , a test can be defined as ‘ any procedure for measuring ability, knowledge and performance’(Richards et al., 1985,p.291) , while Brown (1994, p.252) notes that a test is ‘ method of measuring a person’s ability or knowledge in a given area.’ Brown’s definition seems to be more comprehensive in the sense that it covers all the main components of a test. However what does a test consist of and what is intended for? Firstly a test is a method consisting of a set of techniques , procedures and test items that constitute an instrument of some sort. Secondly, a test has the purpose of measuring the testee’s performance.Thirdly a test is intended to measure a person’s ability or knowledge in a particular area.

According to research studies, tests have another benefit: they make you learn and remember more than you might have otherwise. Although it may seem that all tests are the same, many different types of tests exist and each has a different purpose and style.

## **2.Types of Tests**

### **2.1.Diagnostic Tests**

These tests are used to diagnose how much you know and what you know. They can help a teacher know what needs to be reviewed or reinforced in class. They also enable the student to identify areas of weakness.

### **2.2.Placement Tests**

These tests are used to place students in the appropriate class or level. For example, in language schools, placement tests are used to check a student's language level through grammar, vocabulary, reading comprehension, writing, and speaking questions. After establishing the student's level, the student is placed in the appropriate class to suit his/her needs.

### **3.3.Progress or Achievement Tests**

Achievement or progress tests measure the students' improvement in relation to their syllabus. These tests only contain items which the students have been taught in class. There are two types of progress tests: short-term and long-term.

Short-term progress tests check how well students have understood or learned material covered in specific units or chapters. They enable the teacher to decide if remedial or consolidation work is required.

Long-term progress tests are also called Course Tests because they check the learners' progress over the entire course. They enable the students to judge how well they have progressed. Administratively, they are often the sole basis of decisions to promote to a higher level.

Progress tests can also be structured as quizzes, rather than as tests. They can be answered by teams of students, rather than individuals. They can be formulated as presentations, posters, assignments, or research projects. Structuring progress tests in this way takes into account the multiple intelligences and differing learning styles of the students. Yet many students still expect a "regular test" as a part of "normal learning".

### **4.4.Proficiency Tests**

These tests check learner levels in relation to general standards. They provide a broad picture of knowledge and ability. In English language learning, examples are the TOEFL and IELTS exams, which are mandatory for foreign-language speakers seeking admission to English-speaking universities. In addition, the TOEIC (Test of English for International Communication) checks students' knowledge of Business English, as a prerequisite for employment.

### **5.5.Internal Tests**

Internal tests are those given by the institution where the learner is taking the course. They are often given at the end of a course in the form of a final exam.

### **6.6.External Tests**

External tests are those given by an outside body. Examples are the TOEFL, TOEIC, IELTS, SAT, ACT, LSAT, GRE and GMAT. The exams themselves are the basis for admission to university, job recruitment, or promotion.

### **7.7.Objective Tests**

Objective tests are those that have clear right or wrong answers. Multiple-choice tests fall into this group. Students have to select a pre-determined correct answer from three or four possibilities.

### **8.8.Subjective Tests**

Subjective tests require the marker or examiner to make a subjective judgment regarding the marks deserved. Examples are essay questions and oral interviews. For such tests, it is especially important that both examiner and student are aware of the grading criteria in order to increase their validity.

### **2.9.Combination Tests**

Many tests are a combination of objective and subjective styles. For example, on the TOEFL iBT, the Test of English as a Foreign Language, the reading and listening sections are objective, and the writing and speaking sections are subjective.

### 3.Principles of Writing Good tests

Devising a language test that accounts for the different linguistic variables is not an easy task. Broughton et.al. (1980 ) note that the preparation of tests for educational measurement is time consuming , expensive and requires expertise in statistical techniques as well as in devising suitable tasks for the linguistic assessment to be based on. Additionally, Brown (1994) remark that , ‘ one of the biggest obstacles to overcome in constructing adequate tests is to measure the criterion and not inadvertently something else’. (Brown, 1994, p. 253) In this sense he puts forward three requirements that are axiomatic to qualify a test “good” test , practicality, reliability and validity; in Brown view , if these requirements are carefully met a test can be administrable, dependable and can actually measure what it intends to measure.

Good tests are those that do the job they are designed to do and which convince the people taking and marking them that they work. Good tests also have a positive rather than a negative effect on both students and teachers.

According to Harmer (2007, p. 167) a good test is **valid**. This means that it does what it says it will. In other words, if we say that a certain test is a good measure of a student’s reading ability, then we need to be able to show that this is the case. There is another kind of validity, too, in that when students and teachers see the test, they should think it looks like the real thing - that it has face validity. As they sit in front of their test paper or in front of the screen, the students need to have confidence that this test will work (even if they are nervous about their own abilities). However reliable the test is (see below) face validity demands that the students think it is reliable and valid.

A good test should have marking **reliability**. Not only should it be fairly easy to mark, but anyone marking it should come up with the same result as someone else. However, since different people can (and do) mark differently, there will always be the danger that where tests involve anything other than computer-scorable questions, different results will be given by different markers. For this reason, a test should be designed to minimise the effect of individual marking styles.

When designing tests, one of the things we have to take into account is the **practicality** of the test. We need to work out how long it will take both to sit the test and also to mark it. The test will be worthless if it is so long that no one has the time to do it. In the same way, we have to think of the physical constraints of the test situation. Some speaking tests, especially for international exams, ask not only for an examiner but also for an interlocutor (someone who participates in a conversation with a student). But this is clearly not practical for teachers working on their own.

Tests have a marked washback/backwash effect, whether they are public exams or institution-designed progress or achievement tests. The washback effect occurs when teachers see the form of the test their students are going to have to take and then, as a result, start teaching for the test. For example, they concentrate on teaching the techniques for answering certain types of questions rather than thinking in terms of what language students need to learn in general. This is completely understandable since teachers want as many of their students as possible to pass the test. Indeed, teachers would be careless if they did not introduce their students to the kinds of test items they are likely to encounter in the exam. But this does not mean that teachers should allow such test preparation to dominate their lessons and deflect from their main teaching aims and procedures.

The washback effect has a negative effect on teaching if the test fails to mirror our teaching because then we will be tempted to make our teaching fit the test, rather than the other way round. Many modern public examinations have improved greatly from their more traditional versions, so that they often do reflect contemporary teaching practice. As a result, the washback effect does not have the baleful influence on teaching which we have been discussing.

When we design our own progress and achievement tests, we need to try to ensure that we are not asking students to do things which are completely different from the activities they have taken part in during our lessons. That would clearly be unfair.

Finally, we need to remember that tests have a powerful effect on student motivation. Firstly, students often work a lot harder than normal when there is a test or examination in sight. Secondly, they can be greatly encouraged by success in tests, or, conversely, demotivated by doing badly. For this reason, we may want to try to discourage students from taking public examinations that they are clearly going to fail, and when designing our own progress and achievement tests, we may want to consider the needs of all our students, not just the ones who are doing well. This does not mean writing easy tests, but it does suggest that when writing progress tests especially, we do not want to design the test so that students fail unnecessarily - and are consequently demotivated by the experience.

When designing tests, we can either write discrete items, or ask students to become involved in more integrative language use. Discrete-item testing means only testing one thing at a time (e.g. testing a verb tense or a word), whereas integrative testing means asking students to use a variety of language and skills to complete a task successfully.

A further distinction needs to be made between *direct* and *indirect* test items. A direct test item is one that asks students to do something with language (e.g. write a letter, read and reply to a newspaper article or take part in a conversation). Direct test items are almost always integrative. Indirect test items are those which test the students' knowledge of language rather than getting them to use it. Indirect test items might focus on, say, word collocations or the correct use of modal verbs. Direct test items have more to do with activation, whereas indirect items are more closely related to study- that is the construction of language.

### **Indirect test items**

There are many different ways of testing the students' knowledge of language construction. We will look at three of the most common.

### **Multiple choice**

Multiple-choice questions are those where students are given alternatives to choose from, as in the following example:

#### **Circle the correct answer.**

You *must* \_\_\_\_\_ here on time.

a. to get b. getting c. to have get d. get

Sometimes students are instructed to choose the 'correct' answer (because only one answer is possible), as in the example above. But sometimes, instead, they can be told to choose the 'best' answer (because, although more than one answer is possible, one stands out as the most appropriate), e.g.

#### **Circle the best answer.**

Police are worried about the level of \_\_\_\_\_ crime.

a juvenile b childish c young d infant

Multiple-choice questions have the great advantage of being easy to mark. Answer sheets can be read by computer, or can be marked by putting a transparency over the answer sheet which shows the circled correct letters. Markers do not have to worry, then, about the language in the questions; it is simply a matter of checking the correct letters for each question.

One problem with multiple-choice questions lies in the choice of distractors, that is the three incorrect (or inappropriate) answers. For while it may not be difficult to write one obvious distractor (e.g. answer a 'to get' in the first example above), because that is a mistake that students commonly make, it becomes less easy to come up with three items which will all sort out those students who know how this piece of language works from the ones who don't.

In other words, there is a danger that we will either distract too many students (even those who should get the question right) or too few (in which case the question has not done its job of differentiating students).

Multiple-choice questions can be used to test reading and listening comprehension (we can also use true/false questions for this: students circle 'T' or 'F' next to statements concerning material they have just read or listened to).

The washback effect of multiple-choice questions leads some people to find them unattractive, since training students to be good at multiple-choice questions may not help them to become better language learners. And there is a limit to how much we can test with this kind of indirect item. Nevertheless, multiple-choice questions are very attractive in terms of scorer reliability.

### Fill-in and cloze

This extremely common form of indirect testing involves the examinee writing a word in a gap in a sentence or paragraph, e.g.

Yesterday I went a \_\_\_\_\_ the cinema b \_\_\_\_\_ my friend Clare. I enjoyed the film c \_\_\_\_\_ she did not.

**Gap-fill (or fill-in)** items like this are fairly easy to write, though it is often difficult to leave a gap where only one item is possible. In such cases, we will need to be aware of what different answers we can accept. They also make marking a little more complex, though we can design answer sheets where students only have to write the required word against different letters, e.g.

a \_\_\_\_\_  
b \_\_\_\_\_  
c \_\_\_\_\_

A variation on fill-ins and gap-fills is the cloze procedure, where gaps are put into a text at regular intervals (say every sixth word). As a result, without the test writer having to think about it too much, students are forced to produce a wide range of different words based on everything from collocation to verb formation, etc, as in the following example.

*All around the world, students a \_\_\_\_\_ all ages are learning to b \_\_\_\_\_ English, but their reasons for c \_\_\_\_\_ to study English can differ d \_\_\_\_\_. Some students, of course, only e \_\_\_\_\_ English because it is on f \_\_\_\_\_ curriculum at primary or secondary g \_\_\_\_\_, but for others, studying the h \_\_\_\_\_ reflects some kind of a i \_\_\_\_\_.*

The random selection of gaps (every sixth word) is difficult to use in all circumstances. Sometimes the sixth word will be impossible to guess - or will give rise to far too many alternatives (e.g. gaps c and d above). Most test designers use a form of modified cloze to counteract this situation, trying to adhere to some kind of random distribution (e.g. making every sixth word into a blank), but using their common sense to ensure that students have a chance of filling in the gaps successfully - and thus demonstrating their knowledge of English.

## Transformation

In transformation items students are asked to change the form of words and phrases to show their knowledge of syntax and word grammar. In the following test type they are given a sentence and then asked to produce an equivalent sentence using a given word:

*Rewrite the sentence so that it means the same. Use the word in bold. Could I borrow five pounds, please?*

*l e n d* \_\_\_\_\_

In order to complete the item successfully, the students not only have to know the meaning of borrow and lend, but also how to use them in grammatical constructions.

A variation of this technique is designed to focus more exactly on word grammar. Here, students have to complete lines in a text using the correct form of a given word, e.g.

*It was a terrifying performance.*  
*terrify*

*The acrobats showed \_\_\_\_\_ no fear even though  
absolute  
their feats of \_\_\_\_\_ shocked the crowd into stunned silence.*  
*dare*

These kinds of transformations work very well as a test of the students' underlying knowledge of grammar and vocabulary. However, the items are quite difficult to construct.

There are many other kinds of indirect test items. We can ask students to put jumbled words in order, to make correct sentences and questions. We can ask them to identify and correct mistakes or match the beginnings and ends of sentences. Our choice of test item will depend on which, if any, of these techniques we have used in our teaching since it will always be unfair to give students test items unlike anything they have seen before.

### ***Direct test items***

In direct test items, we ask students to use language to do something, instead of just testing their knowledge of how the language itself works. We might ask our students to write instructions for a simple task (such as using a vending machine or assembling a shelving system) or to give an oral mini-presentation.

There is no real limit to the kinds of tasks we might ask students to perform. The following list gives some possibilities:

### **Reading and listening**

Some reading and writing test items look a bit like indirect items (e.g. when students are given multiple-choice questions about a particular word in a text, for example, or have to answer T/F questions about a particular sentence). But at other times we might ask students to choose the best summary of what they have heard or read. We might ask them to put a set of pictures in order as they read or listen to a story, or complete a phone message form (for a listening task) or fill out a summary form (for a reading task).

Many reading and listening tests are a blend of direct and indirect testing. We can ask students direct language - or text-focused - questions as well as testing their global understanding.

### **Writing**

Direct tests of writing might include getting students to write leaflets based on information supplied in an accompanying text, or having them write compositions, such as narrative and discursive essays. We can ask students to write 'transactional letters'(that is letters replying to an advertisement, or something they have read in the paper, etc). In transactional writing we expect students to include and refer to information they are given.

## Speaking

We can interview students, or we can put them in pairs and ask them to perform a number of tasks. These might include having them discuss the similarities and differences between two pictures ; they might discuss how to furnish a room, or talk about any other topic we select for them. We can ask them to role- play certain situations , such as buying a ticket or asking for information in a shop, or we might ask them to talk about a picture we show them.

When designing direct test items for our students, we need to remember two crucial facts. The first is that, as with indirect tests, direct tests should have items which look like the kind of tasks students have been practising in their lessons. In other words, there is no point in giving students tasks which, because they are unfamiliar, confuse them. The result of this will be that students cannot demonstrate properly how well they can use the language, and this will make the test worthless.

Direct test items are much more difficult to mark than indirect items. This is because our response to a piece of writing or speaking will almost certainly be very subjective -unless we do something to modify this subjectivity.

**Practical Part**  
**Test**  
**Grade VI**

**I. Open the brackets using the right tense form (2 points/0,4 each)**

1. We (to continue) \_\_\_\_\_ the work tomorrow.
2. He (to rise) \_\_\_\_\_ from his seat and went to the door.
3. Yesterday at 7 o'clock we (to watch) \_\_\_\_\_ the new performance.
4. The sun (to rise) \_\_\_\_\_ in the east.
5. While Tom (to speak) \_\_\_\_\_ on the phone, I (to feed) \_\_\_\_\_ my cat.

**II. Make up sentences of the given words. (1 point/0,25 each)**

1. child, The, curious, was, very.

---

2. used, he, when, young, to smoke, was, He.

---

3. foreign, she, What, languages, speak, does?

---

4. both, Canada, will, go, and, I, to, Australia, to.

---

**III. Change the meaning of the sentences using *neither...nor...* (2 points/0,5 each)**

1. Her brother plays both the piano and the flute.

---

2. Both Margaret and Kate are among friends.

---

3. She grew both roses and lilies.

---

4. Both my uncle and aunt go in for swimming.

---

**IV. Fill in the right preposition where necessary. (2 points/0,5 each)**

1. What separates England..... France?

2. The family travelled .....a fast train .

3. The planet we live ..... is the Earth.

4. The children were speaking ..... different languages .

**V. Translate from Armenian into English. (3 points/0,75 each)**

1. Երբ նա վերադարձավ, մենք ջրում էինք այգու ծառերը:

---

2. Աննան գնեց թարմ միս ու ձուկ խնջույքի համար:

---

3. Մեր օրերում մարդիկ ճամփորդում են ինքնաթիռով:

---

4. Դու չե՛ս կարող օգտագործել իմ մեքենան:

---

The test consists of five tasks. The first task is to open the brackets using the right tense form, this task will help to find out how well they have learnt the tense forms. The second task is to make up sentences of the given words, with the help of which I can check the knowledge of Word order. The third task is to change the meaning of the

sentences using *neither.....nor*, which again helps to check the knowledge of Grammatical Rules. The fourth task is to fill in the right prepositions where necessary, which checks the ability of memorizing and using different prepositions. And the last fifth task is to translate from Armenian into English, which gives us the opportunity to check both the knowledge of the vocabulary learnt during the lessons and the grammar.

#### **4. Conclusion**

When we write tests for our classes, we need to bear in mind the characteristics of good tests. We will think very carefully about how practical our tests will be in terms of time (including how long it will take us to mark them).

When writing progress tests, it is important to try to work out what we want to achieve, especially since the students' results in a progress test will have an immediate effect on their motivation. As a consequence, we need to think about how difficult we want the test to be. Is it designed so that only the best students will pass, or should everyone get a good mark? Some test designers, especially for public exams, appear to have an idea of how many students should get a high grade, what percentage of examinees should pass satisfactorily, and what an acceptable failing percentage would look like.

Progress tests should not work like that, however. Their purpose is only to see how well the students have learnt what they have been taught. Our intention, as far as possible, should be to allow the students to show us what they know and can do, not what they don't know and can't do.

When designing tests for our classes, it is helpful to make a list of the things we want to test. This list might include grammar items (e.g. the present continuous) or direct tasks (e.g. sending an email to arrange a meeting). When we have made our lists, we can decide how much importance to give to each item. We can then reflect these different levels of importance either by making specific elements take up most of the time (or space) on the test, or by weighting the marks to reflect the importance of a particular element. In other words, we might give a writing task double the marks of an equivalent indirect test item to reflect our belief in the importance of direct test types.

When we have decided what to include, we write the test. However, it is important that we do not just hand it straight over to the students to take. It will be much more sensible to show the test to colleagues (who frequently notice things we had not thought of) first.

If possible, it is a good idea to try the test out with students of roughly the same level as the ones it is designed for. This will show us if there are any items which are more difficult (or easier) than we thought, and it will highlight any items which are unclear - or which cause unnecessary problems.

Finally, once we have given the test and marked it, we should see if we need to make any changes to it if we are to use some or all of it again.

It is not always necessary to write our own tests, however. Many coursebooks now include test items or test generators which can be used instead of home-grown versions. However, such tests may not take account of the particular situation or learning experiences of our own classes.

## References

1. Alan Davies, Annie Brown, Cathie Elder, Kathryn Hill, Tom Lumley and Tim McNamara (no date) Dictionary of Language Testing. Cambridge University Press.
2. Broughton et al. 1980, Alderson, J. C., Clapham, C. and Wall, D. Language test construction and evaluation. Cambridge: Cambridge University Press.
3. Valette, R. M. (1977). Modern Language Testing, 2nd ed. New York.
4. Jeremy Harmer (2007) "How to Teach English" 2nd Edition