

Caridad Macola Rojo¹
Universidad de Quintana Roo. Campus Chetumal

10 Common Beliefs in Language Testing

Abstract

Testing, being such an important part of the teaching – learning process, is one of the concerns of many teachers who are involved in test design, construction and administration.

In this paper there are 10 topics that most teachers believe or have reflected upon related to testing in general and to the testing of the skills in particular which may serve as basis for future research.

The purpose of this paper is to induce English language teachers to reflect about some beliefs regarding language testing.

The selection of beliefs was derived from: a) an analysis of English language tests and exams recently applied in the University of Quintana Roo, b) discussions carried out in courses on testing given by the author to would be teachers and to actual teachers and c) interviews and surveys made to teachers and students in this University on the subject of “exámenes departamentales”.

The discussion is based mainly on the different types of tests recognized in the literature on the subject.

¹ macola@correo.uqroo.mx

1. Multiple choice exercises serve to test almost everything in language learning

Multiple choice exercises are in the “fixed response format in which a number of possible responses is presented from which the candidate is required to choose”¹ (McNamara, 2000, p.5) Multiple choice is the elicitation technique that is most used in testing not only language but also in many other types of examinations. This form is very widely used because it has many advantages, among which are that they don't require a long time to answer and therefore permit more items to be tested, also that the scoring is reliable and easy to carry out (by non-specialists or machines), and that students do not make any mistakes in writing.

Among the many disadvantages, the salient one is the fact that for the most part they test recognition only. They do not promote critical thinking, there is no possibility of creativity and thus there is a restriction as to what can be tested. With these tests students do not have the opportunity to show that they can communicate, have discussions and perform in the language. “... they are not much used in testing the productive skills of speaking and writing, except indirectly.”² (Ibid, p.6)

Even though objective tests are problem solving in testing a large group of students in a very short period of time, and that multiple choice tests may have *face validity*: “The extent to which a test meets the expectations of those involved in its use, e.g. administrators, teachers, candidates, and test score users; the acceptability of a test to its stakeholders”³ (Ibid, p.133), these tests do not serve to test almost everything in language learning.

The conclusion is then that an objective test cannot really supply information of how much the student knows about a topic or how well he can perform a given task or solve a given problem. “As in other aspects of classroom-based evaluation, **one form of testing** (bolding added) is not necessarily desirable under all circumstances and for all purposes. Rather, judicious use of each form may be called for.”⁴ (Genese, & Upshur, 1997 p.195)

2. A standardized test, made by a group of specialists, is more valid than a test made by a classroom teacher.

This statement would seem true as a test made by a group of specialists with years of experience and tested with groups of students should be more valid than a test made by an individual young and inexperienced classroom teacher. The answer can be found in the difference between standardized and teacher-made tests. In standardized tests the margin of error is minimal, while this is not so in the teacher-made exam.

This is also shown by the difference between proficiency and achievement testing. “Proficiency tests are not based on a particular language programme. They are designed to test the ability of students with different language training backgrounds”⁵ (Alderson, Clapham & Wall, 2001, p.293). A standardized test is usually a proficiency test whose objective is to see whether the student is proficient in English for a given purpose. For example, the TOEFL test is a standardized proficiency test whose objective is to find out whether a person has the skills necessary to be able to register in a postgraduate course in an English speaking country, the United States. One other essential characteristic is that they are not classroom related. The candidates only have to prove that they *know* the English necessary for some objective but not where or how or how long ago this candidate acquired his skills.

Achievement or progress tests, on the other hand, “...are given at the end of the course. The content...is generally based on the course syllabus or the course textbook.”⁶ (Ibid, p.286). These are the tests whose objective is to see what, how much and in what manner the objectives of a given course have been fulfilled. These exams are totally classroom-related and program related.

Upon analyzing *validity* as “the degree to which a test measures what it claims or purports, to be measuring.”⁷ (Brown & Hudson 2002 p.212 from Brown 1996) the conclusion here is that both exams, the standardized proficiency test and the teacher-made achievement test, are *valid* in the extent that each tests what it wants to test; thus, neither is more valid than the other.

“The reader can perhaps take heart that even examination boards do not always do things perfectly; we all have things to learn from relating principles to practice.”⁸ (Alderson et al., p. 3) The person most prepared to know what was taught in the course, how much it was taught and the way it was taught is its teacher or teachers. For this reason, achievement tests have to be teacher made. In other words, specialists, even though they have much more experience in test making than classroom teachers, have no idea of what, how, and how much was taught in a given class, which is the objective of an achievement test.

3. Summative and formative language tests have the same degree of importance in language development as do formal and informal testing.

“Most teachers assess the effectiveness of instruction **informally** on a continuous basis. This is sometimes referred to as **formative evaluation**” (bolding added)⁹ (Genese, & Upshur, p.49) This is a partial test that is applied at the end of a class, a unit or a group of units. It can be frequent or partial and provides much needed information: it tells the teacher how much progress was made and how effective the teaching was, serving as a point of departure for planning the following contents and the methodology. It is essential as it gives the students the opportunity for self assessment, motivating daily studying instead of final exam cramming.

“Summative assessment at the end of the year may be required by the school district or program director in order to assign grades for purposes of certification or promoting students to the next level.”¹⁰ (Ibid) It is an end-of-a-course or year, usually formal comprehensive assessment that is designed to prove that the objectives of the course have been attained. It is a test to sum up classroom activity.

As to the question of formality, informal types of tests are gaining importance in language testing as they provide the teacher with much more detailed information about students’ day to day development of the communication skills. Oral performances, presentations, journals, portfolios, etc. are forms of testing that can portray a student in a much more detailed manner than what can be reflected in a final exam. “Generally speaking, the best evidence of an ESL

student's real, developing language proficiency comes from performance in class. Short in-class "tests" can also yield useful information about students' understanding of both matter and language"¹¹ (Hancock, 1994)

The main difference between the two is that a formative test is done for the purpose of feedback while the summative is not. Neither one is more important than the other. The marks obtained in a final exam are *not* more important in language learning than those marks obtained in frequent evaluations. Each one serves a different purpose. The degree of importance of each type of test is not an issue as they both are important in different ways.

Teachers should consider both types and design a comprehensive evaluation system. In such a well conceived system, the marks given to each type of activity will reflect the whole teaching-learning process. These marks should have a higher percentage allotted to the formative and informal type of testing as in these the teacher and the students can have a frequent appraisal and feedback of each detail of the process; also because in this type it is easier to measure language competence; what, because of its very nature, is not very well revealed in a final comprehensive test.

4. The longer the test the more difficult it is.

Text length is part of the test specifications that test writers must address before undertaking their actual writing. This length is intimately related to *purpose*, which includes the type of test we are going to devise; *contents*, what we are going to test, and the *time* at our disposal, which should not exceed the regular class time.

The discussion here is first based on an analysis of the difference between two concepts which are intimately related in language learning: practice and testing. Practice is done mainly in the classroom or at home, with or without the teacher, in groups or individually, formally or informally in order to attain knowledge or develop skill; testing, on the other hand, is usually carried out in the classroom, with a teacher, individually, and usually formally with the objective of surfacing what, how much and how well learning has taken place.

Testing, then cannot include an example of each of the elements of a grammar exercise, neither can it include all the exercises done in class for writing or speaking or any other of the skills nor all the vocabulary learned in class. In some exams that only test use of English some times there are as many as 100 test items. Apart from the time needed and the risk of actually repeating the items, an exam of such length inevitably will have the same items asked in different ways.

It is obvious that in these cases content validity has not been considered: “Content validity approaches all require the systematic investigation of the degree to which the items on a test, and the resulting scores, are **representative and relevant samples** (bolding added) of whatever content or abilities the test has been designed to measure.”¹² (Brown & Hudson, p.213)

A lengthy language exam far from being a difficult evaluation is very easy because the students can obtain a passing mark only by probability. The idea is to have a representative sample of the items seen in class which includes what is sufficient and essential and eliminates the superficial, additional or repetitious.

5. By applying a productive grammar test we can be sure of the students’ knowledge and performance.

Grammar or use of English is a sub skill in the language learning process and it is of high importance as a basis in the correctness of the production of the language. Reading, writing, listening and speaking are skills that have to be developed in themselves. The distinction that has to be made here is that between tests of skills and subskills or between discrete point and integrative tests. “In discrete point exams, each item tests something very specific such as a preposition or a vocabulary item. Integrative tests are those like dictation that combine various language subskills much the way we do when we communicate in real life.”¹³ (Madsen 1983, p.9)

In many language tests, testing is limited to the subskills of grammar and vocabulary. Baxter (1997) explains this tendency saying:

We substitute the skill that is difficult to teach with one that is easy to teach. While the rules for skills are not very clear, we do have some very good rules for grammar and vocabulary which makes them easier to teach.... So we sometimes believe we are teaching or testing a skill, when really we are practicing or testing grammar or vocabulary.¹⁴ (p.11)

This idea that by testing grammar we can be satisfied that the student knows the language or that a knowledge of grammar will ensure proficiency in language skills is erroneous.“... I recommend tests in which there is less attention paid to specific structure points or lexicon than to the total communicative effect of an utterance.¹⁵ (Brown and Hudson p.21 citing Carroll 1961, p.318).

Knowledge of the rules in isolation, recognition of a given phenomenon in the way a language is structured is an excellent way to learn **about** the language but does not ensure a good performance in its actual use.

6. Speaking and writing tests are subjective tests and their marks must be subjective as well.

This classification of subjective and objective tests strictly speaking does not refer to types of tests but to two different ways for test marking.

Subjective tests ... have the advantage of measuring language skill naturally, almost the way English is used in real life. However, many teachers are not able to score such tests quickly and consistently. By contrast, objective exams...can be scored quickly and consistently.¹⁶ (Madsen, p.8)

A subjective test is one in which the teacher or scorer has to exert a judgment. These are difficult and time consuming. There is not one possible answer and the teacher has to decide how to score his students' abilities as in the tests of speaking and writing. Objective tests, conversely, do not require any judgment on the part of the scorer as they only have one correct

answer, making them possible to be marked by machines or by non specialists. i.e. multiple choice, matching, true/false, etc.

Speaking and writing are skills in which objective tests like multiple choice can only be used in an indirect way but not as a way of measuring student product or performance. The tests for these skills are usually integrative or global as they are task based. Candidates are required to combine many language elements in the completion of a task; then, they are usually tested and marked subjectively.

There are degrees of subjectiveness in scoring and teachers should strive to make their subjective markings as objective as possible. In these cases, the use of marking grids or rater scales -a set of criteria for analyzing a score- where the objectives, the contents and the methodology worked in class in that particular skill are reflected, will reduce considerably the subjectivity in marking the tests.

7. As teachers we cannot bypass a spelling mistake in a writing test.

The writing skill is the Cinderella of language teaching and many a times it comes to our minds only when confronted with having to devise a test and knowing that there should be a writing task in the test. We realize then that we hadn't really developed the ability of writing in our students in that particular period and that we had not given them the means of carrying out the task.

Then we cannot mark functional, sociolinguistic or textual knowledge or the ability of the students to think creatively and develop thoughts, or to select, organize and order relevant information among other abilities.

Then, intent on giving an objective fair mark to the students' papers, we mark grammatical knowledge or formal features of writing like capital letters and punctuation marks, which may have not been taught either but are easier to recognize.

Going back to the definition of an achievement test: an assessment to determine how well the learners have mastered the skills or knowledge presented in the course, this is not always true in designing a writing test: "...designing a test of writing involves defining the ability we are interested in testing for a given test purpose." ¹⁸ (Weigle 2002, p.57)

A spelling mistake, even though highlighted by the teacher, should not be part of the students' grade in a writing test if what we want to test is the ability of the students to write in an appropriate manner for a given purpose with a particular audience in mind. If, on the other hand, the writing task is just a means to show spelling ability as one of the writing objectives, then by all means, spelling should be part of the writing mark.

8. In an oral exam, the students must be recorded so there is an objective proof of what went on during the exam.

Forms of speaking assessment can be varied. In groups, as a whole class activity, in pairs, with the teacher, etc. and these forms depend on the objective or purpose of the evaluation. Many a research has been dedicated to different forms of speaking assessment and ways of marking performance. These studies show that in order to ensure validity and reliability in the assessment of oral competence, it is better that there should be more than one person in the actual assessing.

Different raters, usually other English language teachers, will ensure the objectivity of the results of the test instead of relying solely on the criterion of the students' own teacher. Rating grids, whether holistic "the rating of a performance as a whole"¹⁹ (McNamara, p.133) or analytic "the rating of each aspect of a performance separately"²⁰ (Ibid, p. 131) or a combination of both, compels each rater to concentrate his analysis in the different aspects of the rating grid and can give a very detailed objective analysis of the students' performance. Each teacher has a rater scale developed in advance that contains the criteria used to make the test and the possible marks given to the different performances in order to achieve consistent marking.

Taping the performance of the students is generally practiced by the teacher who doesn't have the possibility of using different raters and rater scales in the assessment of a large group of students with tasks that require more than one student at a time. The drawback in using tapes in an assessment situation is that of negative washback, which does not take place if the students are accustomed to being taped every time that there is a speaking classroom activity. If the students are accustomed to being recorded, this will be an "objective" witness of what went on. If not, to tape the students only during test time will seriously disturb the activity, produce harmful effects on the students and yield unreliable results.

9. Listening assessment should test the ability of understanding details.

Listening is one of the skills that have fewer answers in language learning or assessment. With the advent of the communicative approach and the inclusion of a listening practice in most of the text books, there has been an awareness of the importance of listening in the learning of a foreign language. In general the listening carried out in these text books is done a) to present a new language structure in a "natural" way, b) to familiarize the students with the oral language and give a pronunciation model c) to train the students to obtain *explicit* information from a text. This last one, as it appears more during class practice, is the objective that is usually more tested.

The language teacher is confronted with questions such as: where to get a taped text that is similar in structure to the ones practiced in the classroom? With a similar range of vocabulary? With the same type of pronunciation? The solution is to find audiotexts contained in other text books that have listening tasks similar to the ones that have been carried out in class.

If we consider the skills involved in developing listening in our students,

Listening is a complex process in which the listener takes the incoming data, an acoustic signal, and interprets it based on a wide variety of linguistic and non linguistic knowledge. The linguistic knowledge includes knowledge of phonology, lexis, syntax, semantics, discourse structure, pragmatics and sociolinguistics. The non-linguistic knowledge includes knowledge of the topic, the context and general knowledge about the world and how it works. Comprehension is an on-going process of constructing and interpretation of what the text is about, and then continually modifying that as new information becomes available.²¹ (Buck, p.247)

It is true that a listening assessment should test the ability of understanding details but not the sole ability that should be developed in our classrooms. We can see then that the problem lies in the lack of systematic practicing of different listening abilities. A better knowledge of what listening entails will lead to a more comprehensive practice of this skill with the subsequent enhancement of its assessment.

10. A reading text, in order to be fair, should contain all the words that the students have learned during the period tested.

The skill of reading, as that of listening, in many of the textbook series is developed as a means of presenting a new structure or as a means of reinforcing the structure that has been learned and the tasks are usually focused on understanding explicitly stated information, usually details. With this objective in mind, it is only fair that the vocabulary appearing in our reading tests should be that of the vocabulary that the student has to know because of use in the classroom and because it is in the word lists in the textbooks.

However, the teaching of reading and subsequent testing of the skill is not limited to the objective of understanding information in a text. Some of the objectives of teaching reading: understanding relations within sentences, understanding conceptual meaning, making

inferences, understanding author's intention, skimming, scanning, vocabulary learning strategies, etc. are sometimes not addressed in our classes.

It is not our intention to minimize the importance of vocabulary in a reading text. If a great part of the vocabulary in a given test is not familiar to the students, then it is impossible to obtain any information from that particular text. Nevertheless, some reading tests are in reality vocabulary tests presented in an integrative manner. In these cases the vocabulary in the test needs to have been presented, discussed and learned previously in class. "... the simplest advise is to guard against *only* testing lexical knowledge when attempting to measure reading ability."²²(Alderson, 2001, p.82)

As a conclusion it must be stated that the inclusion of known words in a reading test depends on the testing objectives and on the reading tasks. If the teacher has been working with vocabulary recognition and recall, then the items tested have to be known; if, on the other hand, the teacher has been developing such skills as inferring meaning through contextual clues or through word formation, then the assessment must contain the possibility of showing that the student has mastered this ability.

Conclusions

"The most important consideration in designing and developing a language test is the use for which it is intended, so that the most important quality of a test is its usefulness."²³(Weigle 2002, from Backman & Palmer, p.48)

"Clearly, the alignment of performance standards, instructional objectives and assessment strategies is necessary if student learning and progress are to be evaluated accurately."²⁴(Tambini, 1999)

In order to have good, reliable, valid achievement or progress assessment, the test should reflect the teaching objectives, the contents and the methodology of instruction; in other words, there must be a coincidence between instructional objectives and testing strategies.

References

- Alderson, J. C. 2001: *Assessing Reading*. Cambridge: Cambridge University Press
- Alderson, J. C., Clapham, C. & Wall, D. 2001: *Language Test Construction and Evaluation*. Cambridge: Cambridge University Press
- Baxter, A. 1997: *Evaluating your Students*. London: Richmond Publishing.
- Brown, J. D. & Hudson, Th. 2002: *Criterion-referenced Language Testing*. Cambridge: Cambridge University Press
- Buck, G. 2001. *Assessing Listening*. Cambridge: Cambridge University Press
- Genesee, F. & Upshur, J. A. 1997: *Classroom-Based Evaluation in Second Language Education*. Cambridge: Cambridge University Press
- Hancock, C. 1994. Alternative Assessment and Second Language Study: What and Why? *L & L Digest July 1994*. Retrieved September 21, 2005 from Center for Applied Linguistics <http://www.cal.org/>
- Madsen, H. 1983: *Techniques in Testing*. New York: Oxford University Press
- McNamara, T. 2000: *Language Testing*. Oxford: Oxford University Press
- Tambini, R.F. 1999: Aligning Learning Activities and Assessment Strategies in the ESL classroom. *The Internet TESL Journal*, Vol V, No. 9, September 1999. Retrieved September 21, 2005, from <http://iteslj.org/Articles/Tambini-Aligning.html>
- Weigle, S. C. 2002: *Assessing Writing*. Cambridge: Cambridge University Press

Biodata

Caridad Macola Rojo

BA in the Teaching of English from the Havana Business University and Doctor in Education from the University of Havana. Full professor in the Language and Education Department in the University of Quintana Roo. Has written in co-authorship textbooks for the teaching of English for specific purposes, Readings for Social Sciences I and II; also, textbooks for the teaching of translation: Translation I and II and English Grammar Workbook for the subject of Contrastive Grammar. Her most recent publications are methodological guides: English Grammar I, English Grammar II and Translation I.