

THINKING ABOUT TESTING

Anna Maria Becker Maciel *
Fátima Moreira Harbich *
Lourdes Marmet *
Maria Helena Curcio Celia *

1. INTRODUCTION¹

This paper originated in a request to organize a Special Interest Study Group on Testing at the IV SEMPUI – *Seminário Nacional de Professores Universitários de Inglês*, held in Florianópolis in July 1982. The issue is likely to raise interest whether of teachers of English as a foreign language (EFL) or of any person involved with the learning-teaching process. There is a perpetual claim against tests. Neither testers nor testees are satisfied with the content, the purpose or the manner of most tests available or teacher-made.

A wide gap between the recent developments applied in EFL classes in terms of teaching methodologies and testing resources is apparent. Indeed, EFL teachers seem quick to incorporate new insights in their class learning activities but do not dare to change their old testing techniques dating from the 60's and even before. Not surprisingly testing procedures do not match the teaching-learning ones.

The present work aims at a reconsideration of the ordinary testing methodology on the light of the principles put forward by the communicative approach to language teaching. A brief outline of the state of the art is offered placing the test in a historical perspective. An article by Keith Morrow is studied, not only to explain what has already been covered but also to provide a basis for further discussion; a review of this article is included in the present paper.

* Professoras da Universidade Federal do Rio Grande do Sul (UFRGS), Departamento de Letras.

¹ The authors are members of a research team engaged in the *Projeto Nacional de Inglês Instrumental em Universidades Brasileiras*, coordinated by PUC/SP and Prof. Nora Ther Thielen at UFRGS.

Finally some suggestions for communicative test items are presented. The appendices include a workshop on test items, a glossary and an annotated bibliography.

2. THE STATE OF THE ART

2.1. Tests and the teaching-learning process

Whenever we deal with testing, the concepts of measurement and evaluation come to our minds. The difference between them lies in the fact that the former provides data from which the latter is inferred. The evaluation process consists of a series of inferences and relationships which are obtained from the construction and application of tests. Therefore, a test is useless if its results do not allow for the evaluation process.

Tests are tools used to measure and evaluate the learner's growth and development (Bertrand & Cebula, 1980:1). The test is not an end in itself, it is a mean to achieve a more efficient teaching-learning process. Its role is to gather data which are used by the teacher with the purpose of giving the learner what he really needs in terms of classroom work, course design and materials writing.

The test is thought to be the procedure that naturally follows pedagogic-didactic activities. Its construction, however, must be considered before the effective development of the curriculum has taken place. This is equivalent to say that the test establishes the guidelines for course design. Such a view requires the test to be seen from three different perspectives: the learner, the teacher and the course.

Even knowing that his ultimate goal is not to pass the exam, the student really studies in order to get good grades. Tests and examination are therefore considered to be sorts of guidelines not only of each semester but also of the whole course. Moreover, the learner studies according to the way in which he is going to be assessed. If the test requires a personal and creative use of what has been learnt, the student will try to develop critical judgement. If the test rests upon memorization, he will study in a mechanical way. The tests, then, give the students a functional definition of the course objectives, guiding and stimulating their endeavour.

As far as the teacher is concerned, the making of any test is the opportunity to review objectives and procedures so that constant feedback over the instructional process can be established. If the teacher keeps the tests in mind it will be easier to refer to the objectives in a behavioural way, in the sense of operational terms.

Considering the course, tests are made and administered according to the instructional process. *Instruction takes place whenever there is a process for producing planned changes in the behaviour of the students.* This involves three basic steps: 1. determining what is going to be learnt; 2. carrying out the actual instruction; 3. evaluating the change in the learner (Lindvall & Nitko, 1975:10-1). The latter is the only way of knowing whether instruction has taken place or not. The test, thus, is closely related to the activities developed in the classroom.

To gather the learners' outcomes in order to organize scores statistically is not the only objective of the test. From this point, evidence should be gathered on the effects of the teaching-learning process. These effects are the changes in the learner's learning process. These effects are the changes in the learner's behaviour, caused by the activities developed during the course. In addition, it is up to the teacher to decide the content, the purpose and the procedures to be used when testing.

Evaluation provides data for the instructional process to be established in a cyclic manner, on condition that essential principles be followed. Four principles may be emphasized:

1. to define the objectives of the curriculum;
2. to place these objectives in terms of behaviour;

3. to create situations that elicit the development of such behaviour;
4. to evaluate the input and the output behaviours.

Since the test is the instrument from which we get measurement to evaluate the whole teaching/learning process, there must be coherence concerning several aspects, namely, underlying linguistic theory, underlying learning theory, course objectives, teaching methodology and testing approach. It must be emphasized that all these aspects depend on choice: we choose theories, methods, objectives and also choose the sample of the language we are going to teach and test.

One factor that has to be taken into account when deciding what and how to teach and test is whether that item can be systematically described. Another factor to be considered is why we are teaching and testing that item; it is related to the choice of our objectives and knowledge of the learner's needs.

The most important aspects involved in assessing progress in learning a foreign language are how it works and how the learner will use it. A test is "any observable activity a student is asked to perform under controlled conditions in order to determine his capacity to perform similar activities under less rigid controls" (Oller, 1973:184).

In order to assess the learner's performance and to infer other results related to the whole teaching/learning process, we must be able to answer an important question:

DOES THE INSTRUMENT WORK?

This question may be interpreted in at least four different ways:

- a) Does it measure consistently?
- b) Does it distinguish between one pupil and another?
- c) Does it measure what it is supposed to measure?
- d) Does it consist of readily usable material?

These four questions refer to RELIABILITY, discrimination, VALIDITY, practicality.

2.2. Historical Perspective

As testing is part of the language teaching/learning process, it has always reflected ideas, procedures and theories about this process. This can be better explained providing an outline of the general trends in language teaching. If we take a look at the historical development of language teaching, from before World War I to our days, it can be seen that it has always been a matter of interdisciplinary interest.

Since at least three elements provide the basis for language teaching – the language, the learner, the objectives – a number of disciplines must play different roles. Linguistics provides a linguistic theory. Psychology is going to take care of the learner as an individual, whereas sociology studies the learner while a human being in social interaction. Pedagogy contributes for the process in the classroom, through methodology and techniques. As a consequence of concentrating on the same subject, these areas have resulted in new disciplines, such as applied linguistics, psycholinguistics and sociolinguistics.

For the sake of brevity and in order to have a didactic view of the teaching of English as a second or foreign language, let us consider three basic stages. It is important to keep in mind that these stages have not been developed in extreme and separate ways; their features sometimes overlap. These three stages have been named in different ways in the development of language testing in this century. Spolsky (1975) has used the labels 'pre-scientific', 'psychometric-structuralist', 'psycholinguistic-sociolinguistic'; Morrow (1979) has characterized these stages as the 'Garden of Eden', the 'Vale of Tears' and the 'Promised Land'.

Let us first specify the three stages in terms of the general view of language teaching they adopt:

1) TRADITIONAL:

- linguistic theory – traditionalist
- no psycho-sociological theory involved
- pedagogical implications – grammar-translation method

2) ANALYTICAL:

- linguistic theories – structuralism/TGG
- psychological theories – behaviourism/cognitivism
- pedagogical implications – audiolingual method/cognitive approach

3) INTEGRATIVE:

- linguistic theories – semantics, pragmatics, register, speech act theory, discourse analysis
- psycho-sociolinguistic aspect – communication, needs analysis
- pedagogical implications – notional/functional syllabuses

Parallel to this progress in language teaching research, there should have been a similar move in terms of testing. But the relationship is not perfect. We can say that there is a gap between progress in language teaching approaches and improvement in testing procedures.

We are going to concentrate on stages (2) and (3) to analyse the developments in testing. Lado's legacy is the point of departure for any discussion on this area. In testing as in teaching there seems to be a tension between the analytical and the integrative points of view. Thus, Davies' dichotomy ANALYSIS vs INTEGRATION provides some data for this comparison (1978:150):

analytical	integrative
discrete-point	wholistic
linguistic competence	communicative competence
form	function
usage	use
norm-referenced	criterion-referenced
reception	production
summative	formative
idealisation	raw data
deep structure	surface structure
reliability	validity

In theoretical terms, it is a characteristic of stage (2) that tests should consist of discrete-point items. This view considers that it is important to base all teaching and testing on the results of a contrastive analysis. Ideally, testing would be a totally objective matter, since this atomistic approach emphasizes quantifiable data. Influences of the structuralist view of language and the behaviourist view of learning can also be seen in the adoption of norm-referenced criteria. However, all these principles have never appeared in such an extreme form in practice.

When we move to stage (3), it becomes apparent that communicative views of language teaching do play their roles. The problem has already been stated that evaluation has hardly been looked at, either in terms of assessment of the communicative abilities of the learner or the efficacy of the programme he is following. The consequence is that syllabuses aimed at developing communicative interaction are often followed by inadequate evaluation instruments. There is a difficult connection to be made between evaluation, variable learner characteristics

and a psycho-socio-linguistic perspective on 'doing' language-based tasks. Nowadays, there is a greater interest in language in context and in social interaction.

2.3. Modern Trends

Modern trends in language teaching focus language in use. The wholistic aspect of the communicative event is emphasized, the integrative approach to testing is valued rather than the discrete view point. To place the testee in authentic situations is the major concern of the tester.

Of course we are not going to provide a completely authentic situation, for example, in language tests; just because they are language tests, they are not real communication (Alderson, 1981:3). Nevertheless, some criteria may be adopted in the pursuit of communicative tests. We have already seen that we cannot test what is not systematic and language in use is not systematic concerning individuals and different points in time. Usage is systematisable, particularly grammar, syntax. However, linguistic competence is only a part of communicative competence.

Morrow (1979:149-50) lists some of the features of language use that lack measurement in conventional tests:

- language in use is interaction-based;
- it has unpredictable data;
- any use of language takes place in a context, with a purpose;
- its assessment must be made in terms of the effective performance of the learner in an authentic situation;
- it is judged on the basis of behavioural outcomes.

Since it is impossible to reproduce all the characteristics of real language used communicatively, according to Carroll (1980:13-6), a good test will show an optimum balance of these four characteristics (RACE):

- **Relevance:** how relevant is the behaviour being tested to the meeting of communicative needs?
- **Acceptability:** will the users of the test accept its content and format?
- **Comparability:** can the test scores obtained at different times and from different groups be compared?
- **Economy:** do the tests provide as much information as is required with the minimum expenditure of time, effort and resources?

All of these characteristics together with the question **WHAT CAN THIS CANDIDATE DO?** lead us to tests of communicative performance. Although we cannot deny that the learner connects what goes on in the language classroom with what goes on in the real world, they also bring about problems in terms of extrapolation and assessment.

The first problem – **EXTRAPOLATION** – can be dealt with according to the following process: since the language curriculum is developed on the basis of the communicative needs of the learner, his communicative performance is going to be assessed; but at the same time, the application of enabling skills requires creativity on the part of the learner. This creativity reflects his competence, because the use of a language is the objective, and the mastery of the formal patterns, or usage, of that language is a mean to achieve this objective. Equipped with a mastery of these language patterns, it is hoped that the user will learn how to cope with the situations he finds himself in. From the 'use' point of view, language loses its appearance of unity depending on different patterns of communication, and must be taught and tested according to the specific needs of the learner.

As far as the second problem – **ASSESSMENT** – is concerned, a solution can also be

found. Carroll (1980:31) suggests criteria to be followed in order to guarantee that the learner is going to perform a communicative task. Provided that the objectives are clearly stated, the teacher is supposed to know exactly what kind of output he expects from the learner. This means that the characteristics of the language being taught and tested are well established and that the teacher can describe the sort of knowledge the learner needs in order to use that type of language in that particular context, with that specific communicative purpose in mind. It is not denied that the construction of communicative performance tests according to these criteria presents subjectivity; but performance scales are suggested to provide a basis for this issue.

3. COMMUNICATIVE LANGUAGE TESTING: REVOLUTION OR EVOLUTION

Review of an article by Keith Morrow

MORROW, K. Communicative Language testing: revolution or evolution. In: BRUMFIT, C.J. & JOHNSON, K. *The communicative approach to language teaching*. London, OUP, 1979, p. 143-57.

The aim of this paper is to consider the design of foreign language tests under the light of the communicative approach. The author starts by quoting the hope of Wilkins (1976:82) that the introduction of the notional syllabus should trigger the development of parallel testing procedure and by regretting the inadequacy of the available evaluating techniques to incorporate the recent linguistic and pedagogic developments.

Looking back to a historical perspective of language evaluation (Spolsky, 1975), he characterizes the three different stages of the testing methodology as the Garden of Eden (the pre-scientific stage), the Vale of Tears (Lado's psychometric-structuralist approach) and the Promised Land (the psycholinguistic-sociolinguistic view). He disregards the first one as gone for ever but he lingers on the next stage on the grounds of its pervading influence. Finally, he offers for further consideration insights on a communicative test design and discusses the most relevant issues involved.

Morrow calls into question the features claimed by Lado as the essential ones to test design, namely, reliability and validity. The former, deeply grounded on objectivity, and the latter, derived both from the atomistic nature of language and the behavioural approach to language learning, characterizes the Vale of Tears.

Lado's objectivity is challenged in terms of selection of samples of language.

A high degree of subjectivity is detected in the choice of items picked up from the examiner's own system of language which is offered to the examinee as a model to be conformed. Lado's objectivity seems to lie mostly on the scoring system and not on the actual construction of tests.

Moreover, the author advocates the dynamic concept of transitional competence (Corder, 1973) which enables the learner to produce and use an interlanguage (Selinker, 1972). These concepts allow the tests to show how far and how well the learner has moved towards an approximation of a native speaker's system of language. In this way, the quality of the language mastered is assessed rather than the quantifiable amount of language acquired.

Morrow is cautious about validity: a test is valid whenever the criteria on which it is based are thought to be true; once they are questionable, its validity is doubtful. In Lado's approach, the challenged criterion is the assumption that the knowledge of isolated segments of language tested in discrete items accounts for the ability of combining them in new and appropriate ways to meet new and appropriate linguistic demands.

As a counterpart to Lado's ideal pure discrete items, Morrow presents the attempts of Carroll (1968) followed by those of Oller (1971, 1973). Both developed integrative test items trying to mobilize linguistic competence and performance ability in ordinary situations of

language use. They encompass basic language processing mechanisms and a wide range of structural and lexical items in meaningful contexts. Notwithstanding, Morrow does not believe that they have succeeded in creating test situations which may parallel real life ones.

The features of language in use in genuine real life situations that conventional tests, whether discrete point or integrative, fail to measure are expected to be taken into account by communicative tests. The communicative tests belong to the phase Morrow calls The Promised Land. He identifies their main features as follows:

Interaction-based is the quality of a language event that demands the presence at least virtual, of an addresser and an addressee whose expectations will monitor both the content of the message and the way in which it is expressed.

Unpredictability is the quality of an interaction that cannot be planned beforehand and where the participants must process unknown data in real time.

Context concerns the circumstances, linguistic or not, within which any language event takes place. While the linguistic context relates to the actual system of the language, such as textual cohesion, the context of situation relates to physical environment, role and status of participants, attitude and formality.

Purpose is the feature that allows the participants to identify the goals each has in mind and to encode appropriately their goals.

Performance means how the participants put language in use.

Authenticity is the quality of a language sample which is genuine, not simplified.

Behaviour based is the feature that concerns what each participant can actually achieve through language.

The extent to which a communicative test does the job for which it is used raises a number of problems in terms of extrapolation, assessment and content. These problems challenge the conventional concepts of validity and reliability. Notwithstanding, Morrow believes that to design a valid reliable communicative test is a feasible and worth-doing task.

The main question concerning extrapolation is how to ensure that the candidate's achievement of a given task entails the ability to perform other similar tasks. As an operational solution, Morrow proposes to break down the performance of a global communicative task into small subtasks. A model for such a breaking is implicit in Munby (1978) and has been implemented for testing purposes by B.J. Carroll (1978).

Carroll's model shows for any given task the enabling skills which have to be mobilized to complete it. It should be assumed that the candidate that performs well in individual enabling skills will be able to use them in a global real situation. Nevertheless, the limitations concerning the relationships of the whole and the parts remain.

Performance cannot be assessed in quantifiable terms. Though the assessment might be converted into numerical scores, it is primarily qualitative. Qualitative criteria risks to be subjective. Morrow suggests the use of an operational scale of attainment developed by B.J. Carroll (1978). Carroll takes into account ten performance criteria in order to confront the testee with a communicative task, relating either to the task or to the testee's performance. Morrow considers such a scale an alternative way of assessing the quality of performance at different levels combining face validity and, at least, potential reliability.

Content is, of course, the central issue of a communicative test that claims to be performance-based. In communicative terms there is no single overall test of language proficiency. There

are tests of proficiency at different levels in terms of specified communicative criteria. These criteria cannot be ready made, they depend on the communicative needs of the learner.

The difference of proficiency levels and the specification of communicative criteria challenge the traditional concept of pass and fail. Even the low scorer can be shown to have achieved. Different scores can be related to specific communicative objectives. Moreover, the candidate should be aware of the types of operation, the content areas and the assessment criteria to be adopted.

In short, Morrow states that the designer of a communicative test should start by a deep consideration of what sorts of things people actually use language for in the areas in which he is interested in. Then he should answer the following questions:

1. What are the performance operations I wish to test?
2. What level of proficiency will I expect?
3. What enabling skills are involved? Do I wish to test control of these separately?
4. What content areas in terms of types of operation and types of text are appropriate?
5. What format will I adopt? It must allow for both reliability and face validity as a test of language use.

4. SOME SUGGESTIONS FOR COMMUNICATIVE TEST ITEMS

In order to check the student's communicative performance we need to devise tests which actually measure his capacity to communicate. A structurally competent learner can be communicatively incompetent. He may have developed the ability to produce grammatically correct sentences, yet he is unable to perform a simple communicative task, such as, making a phone call or asking for information.

A communicative test is one which is based not only on a selection of items chosen on linguistic grounds alone, but on the learner's needs and requirements to use the language. What is important in such a test is the communicative effectiveness rather than the formal correctness. In this way tests can be pragmatically or non-pragmatically valid.

In pragmatically valid tests the examinee must do more than the mere recognition of language items; in fact, he will be using language in a situation similar to real life. For example, if I use an item such as: "In the following list tick all the words which rhyme: sit - seat - feet - sheet" (Davies, 1979), the ability to distinguish sounds is being used only for the sake of it and not to perform a communicative task. Whereas in a dictation, for instance, the testee must be able to understand the meaning of what is being dictated so as to write it appropriately. There are several situations in real life where dictation takes place: dictation of letters, addresses, telephone messages, instructions, notes, etc. Dictation, therefore, can be considered a communicative procedure.

Oller (1979: 267) suggests that dictation can be used in varied forms:

- a) standard dictation;
- b) partial dictation, one which combines dictation with cloze procedure: some portions of the text are deleted and the examinee must fill in the missing parts while listening to the complete version;
- c) dictation with some kind of background noise;
- d) dictation-composition, one in which the examinee listens to a passage, takes down some notes and then tries to rewrite it.

Another valid form of testing is the cloze technique. There are always situations where one must have to read a letter with illegible handwriting or a photocopy whose printing has not come out very well or listen to people talking in a noisy place, or understand the words of a

song. The procedure consists of deleting every 5th, 6th, 7th, etc. word of a text and making the student fill in the gaps with appropriate words, such as in the following example:

It is true that persons . . . view the treatment of mental . . . from a clinical perspective tend . . . explain socioeconomic and ethnic differences . . . biological terms (Oller, 1979: 341).

The cloze procedure involves the abilities of guessing, of using redundancy and applying background information. They are all valuable skills to be developed through this procedure. While reading an incomplete text the student will be using them and the better he can do it, the more likely he is to succeed in his task.

There are some aspects which must be taken into account while preparing a pragmatically valid item:

- authenticity of material;
- contextualization: who is communicating to whom, where and when;
- purpose: why and what for;
- information gap: testees should be allowed to have a free choice on what they are going to say, pictures should be unknown to them, information should be new.

Let us see some pragmatically valid tasks involving the four skills:

a) *writing activities:*

- letter writing (replies, applications for jobs, for courses, magazine subscriptions, personal letters about one's holidays, social affairs, etc.);
- summaries;
- paraphrases (provide a much simpler version of a passage);
- narratives (first day at school);
- professional reports;
- note-taking;

b) *reading activities:*

- proof reading;
- reading for enjoyment;
- reading for information: telephone directory, dictionaries, encyclopaedias, manuals, guide books, travel brochures, newspapers, magazines, journals, etc.;
- reading aloud: radio broadcasts, stories for children, professional purposes, dramatic reading;

c) *speaking and listening activities:*

- oral interview: narrating previous job experience, holiday travels, answering the Customs officer, etc.;
- making and answering a phone call;
- asking for information (how to get to the post-office, etc.);
- discussing one's point of view on a variety of topics (divorce, nuclear weapons, etc.);
- describing the education system of one's country, etc.

Finally, let us show some examples of communicative test items involving the four skills.

LISTENING

Today is 10th July. You live in London. Tomorrow you plan to visit Brighton on the south coast. Listen to the weather forecast and decide if you should take your raincoat with you to Brighton.

(Geddes, 1981:81)

SPEAKING

- a) You're going down the road to buy some cigarettes. Offer to buy your father the newspaper at the same time.
- b) A friend of yours is going away for the weekend. Offer to water his or her plants and feed the cat.

(Lake, 1979:43)

WRITING/READING

You have been asked by some friends from your country to book a four day coach tour for them for the Easter weekend. There are six of them, and they wish to visit some places of scenic beauty and cultural significance. They can afford up to £ 70.00 each.

- a) Read the texts which describe a number of different tours.
- b) Make a list of four suitable tours.
- c) Write a letter to the tour operator or operators to find if there are any places available on the tours concerned.

(White, 1981: 89)

WRITING

- a) Using ideas from the paragraph on memories of youth above write a letter to your boyfriend/girlfriend telling about your childhood, where you used to live, what you used to do, etc. Try to use 'used to' and 'would'.
- b) In Winton there is a very good sports club which welcomes applications from overseas visitors.
Arturo Catania applied to join this club. He wrote a letter to the club secretary who used it to fill out an application form for Arturo. Here is the form. Use it to write Arturo's letter of application.

WINTON SPORTS CLUB: MEMBERSHIP APPLICATION

Surname: *CATANIA*
Date of birth: *4.8.46*
Marital status: *M*

First name(s): *ARTURO*
Nationality: *Italian*

Address: *29, Gosforth Close, Hamford, Nottingham - NG 167 EA*
Telephone number: *(0273) 51469*

Occupation: *Doctor*

(Johnson, 1981: 96)

5. CONCLUSION

MORROW (1979: 143) himself realizes that this issue — TESTING — is a polemical one. Conventional testing procedures that were thought to be reliable and valid are challenged in favour of a new approach to test design. This novel perspective focuses on communicative principles of teaching and testing. Nevertheless this view still presents limitations and drawbacks concerning the application of communicative criteria to the evaluation process.

It is impossible to offer ready-made solutions, but teachers and research workers may feel optimistic and encouraged to attempt the development of up-to-date testing methodology derived from a view of language as communication. What is crucial is the awareness that there is a gap between the communicative teaching methodology and the conventional testing procedures available and that something must be done in order to change this state of affairs.

BIBLIOGRAPHY

- ALEXANDER, L.G. (1976) Question and answer: graded oral comprehension exercises, apud PAULSTON, C.B. & BRUDER, M.M. *Teaching English as a 2nd language: techniques and procedures*. Massachusetts, Winthrop.
- ALDERSON, J.C. (1981) *Communicative language testing*. A personal account of the discussion by J. Charles Alderson. Lancaster, mimeo.
- ARCHER, N.R., NOLAN-WOODS, E. (1977) *Practice tests for proficiency 1st series*. Middlesex, Nelson.
- BELL, R.T. (1981) *An introduction to applied linguistics: approaches and methods in language teaching*. London, Batsford Academic and Educational.
- BERTRAND, A. & CEBULA, J.P. (1980) *Tests, measurement and evaluation*. A developmental approach. Reading, Massachusetts, Addison-Wesley.
- CANDLIN, C. (1981) (ed.) *The communicative teaching of English*. Principles and an exercise typology. London, Longman.
- CARROL, B.J. (1968) The psychology of language testing. In: DAVIES, Alan (ed.). *Language testing Symposium*. A psycholinguistic approach. London, Oxford University Press, p. 46-69.
- _____ (1978) *An English Language testing service: Specifications*. London, British Council.
- _____ (1980) *Testing communicative performance*. An interim study. Oxford, Pergamon.
- COLLINS, H.W. (1976) *Educational measurement and evaluation: a worktext*. 2 ed. Glenview, Scott, Foresman and Company.
- CORDER, S.P. (1973) *Introducing applied linguistics*. Harmondsworth, Middlesex, Penguin.
- DAVIES, A. (1979). Language Testing. Parts 1 & 2. Survey Articles. In: *Language Teaching and Linguistics*. Abstracts. Cambridge, Cambridge University Press, p. 145-59, 215-31.
- ETHERTON, A.R.B. (1969) *Objective English tests*. London, Longman.
- GEDDES, M. (1981) Listening. In: JOHNSON, Keith & MORROW, Keith (ed.). *Communication in the classroom*. Applications and methods for a communicative approach. London, Longman, p. 78-86.
- GOOD, C. (1959) (ed.) *Dictionary of education*. New York, McGraw-Hill.
- HARRIS, T.L. & HODGES, R.E. (1981) (ed.) *A dictionary of reading and related terms*. Newark, IRA.
- HEATON, J.B. (1975) *Writing English language tests: a practical guide for teachers of English as a second or foreign language*. London, Longman.
- HYMES, D.H. (1979) On communicative competence. In: BRUMFIT, C.J. & JOHNSON, K. *The communicative approach to language teaching*. London, OUP, p. 2-26.
- INGRAM, E. (1977) The construction of language tests. In: ALLEN J. & DAVIES, A. *Testing and experimental methods*. Edinburgh Course in Applied Linguistics, V. 4, London, OUP.

- INGRAM, E. (1977) Basic concepts in testing. In: ALLEN, J.P.B. & DAVIES, A. (ed.). *Testing and experimental methods*. Edinburgh Course in Applied Linguistics, V. 4, London, OUP.
- JOHNSON, K. Writing. (1981) In: JOHNSON, Keith & MORROW, Keith (ed.). *Communication in the classroom*. Applications and methods for a communicative approach. London, Longman, p. 93-107.
- LADO, R. (1967) *Language testing*. The construction and use of foreign language tests. London, Longman.
- LAKE, S. (1979) *Intermediate Context English*. London, Macmillan.
- LINDVALL, C.M. & NITKO, A.J. (1975) *Measuring pupil achievement and aptitude*. 2 ed. New York, Harcourt Brace Jovanovich.
- MORROW, K. (1979) Communicative language testing: Revolution or evolution? In: BRUMFIT, C.J. & JOHNSON, K. (ed.) *The communicative approach to language teaching*. London, Oxford University Press. p. 143-57.
- MUNBY, J. (1978) *Communicative syllabus design*. A sociolinguistic model for defining the content of purpose - specific language programmes. Cambridge, Cambridge University Press.
- OLLER, J. (1971) Dictation as a device for testing foreign language proficiency. *English Language Teaching Journal*. 25, 3.
- _____ (1973) Discrete - point test versus tests of integrative skills. In: _____ & RICHARDS (ed.). *Focus on the learner: pragmatic perspectives for the language teacher*. Rowley, Massachusetts, Newbury House, p. 194-99.
- _____ (1979) *Language tests at school: a pragmatic approach*. London, Loryman.
- _____ (1980) Language testing today: an interview with John Oller. In: BOUCHARD, D.L. & SPAVENTA, L.J. (comp.). *A TEFL anthology: selected articles from English teaching Forum 1973- 78*. Washington, English Teaching Division.
- REMMERS, H. (1965) *A practical introduction to measurement and evaluation*. 2 ed. New York, Harper & Row.
- SELINKER, L. (1972) Interlanguage. *International Review of Applied Linguistics*. 10, 3.
- SPOLSKY, B. (1975) Language testing: art or science? *Address to the Fourth AILA Congress*, Stuttgart. Summarized in Valette 1977.
- WHITE, R.V. (1981) Reading. In: JOHNSON, Keith & MORROW, Keith (ed.). *Communication in the classroom*. Applications and methods for a communicative approach. London, Longman. p. 87-92.
- WIDDOWSON, H.G. (1978) *Teaching language as communication*. London, OUP.
- WILKINS, D.A. (1976) *Notional Syllabuses*. A taxonomy and its relevance to foreign language curriculum development. Oxford, Oxford University Press.

APPENDIX 1

Workshop

This workshop was carried out at the Special Interest Study group on Testing in IV SEMPUI, 1982. The participants formed four groups and each group had to analyse the five features of language use concerning all of the six items that appear on the table below. The lecturers tutored the group work and after twenty minutes the whole class gathered in order to discuss the results.

The test items presented the following features:

	1	2	3	4	5	6
content				X		X
purpose				X		
authenticity				X		X
behaviour-based				X		
interaction-based				X		X

TEST ITEM ONE

This exercise is to discover what English sounds you can hear clearly. Answer every question. I am going to read three words. Check the numbers of the words that are the same. When no two words are the same, check the zero.

Example: A) hat - hat - hat 1-2-3 0

B) beat - hit - hit 1 2-3 0

C) but - hot - that 1 2 3 0

The words will be read only once. Blank are counted as wrong items.

A) 1 2 3 0

B) 1 2 3 0

C) 1 2 3 0

D) 1 2 3 0

E) 1 2 3 0

This test should be administered by a native speaker of standard American English or by one who can speak English with all the pronunciation contrasts of a native speaker. The examiner reads in a clear natural style the following words:

A) cat cat cot

B) run sun run

C) last last last

D) beast best best

E) pair fair chair (based on Lado, 1967:73)

TEST ITEM TWO

Instruções: Este é um teste de compreensão da língua falada. Você ouvirá um texto em inglês e responderá em português a perguntas feitas em português. Leia, agora, atentamente, as perguntas e concentre sua atenção nas respectivas respostas ao ouvir o texto. Você ouvirá o texto duas vezes. Cada resposta correta vale 1 ponto. Escreva a resposta nas linhas pontilhadas.

Perguntas:

- 1) Qual a profissão do Robert Jones?
- 2) Qual a profissão de Miss Green?
- 3) Onde estão eles no momento?
- 4) O que Jones pede a Miss Green?
- 5) O que ela lhe oferece?
- 6) O que Jones pergunta a Miss Green?
- 7) Por que ela não pode responder à pergunta?

The students listen to the following text:

Inspector Robert Jones is a detective. He lives near a library. Inspector Jones is in the library now. He is talking to Miss Green. Miss Green is the librarian.

- Good morning, Miss Green.
- Good morning, Inspector. Can I help you?
- Yes, you can. I'm looking for a good book.
- This is a good book, Inspector.
- What is it?
- A detective story.
- Good. I like detective stories. What's the murderer's name?
- I can't tell you that, Inspector. I'm a librarian. I'm not a detective. (adapted from Alexander, L.G. 1976)

TEST ITEM THREE

Choose the verb form that can be used in the sentence.

- a) We caught the boys the fruit.
1. take 2. took 3. taken 4. taking 5. takes
- b) He rushed out of the room, the door as he went.
1. slam 2. slams 3. slammed 4. slamming 5. to slam
- c) I didn't know he already.
1. finish 2. had finished 3. have finished 4. has finished 5. finishing
- d) I know he the driving test if he tried hard enough.
1. will pass 2. can pass 3. has passed 4. could pass 5. passing

(from: Etherton, A.R.B. 1969:34)

TEST ITEM FOUR – WRITTEN TEST

These are some situations you could find yourself in. Someone is asking you why you did (or did not do) something. Give your reply and reason why (not).

Teacher: – You are late again. Couldn't you be on time for once?

You: – No, I couldn't because

Teacher: – Couldn't your father bring you in his car?

You: – No, he couldn't because

(from: Candlin, C. 1981:174)

TEST ITEM FIVE – ORAL TEST

Look at this picture carefully and be prepared to answer some questions about it.



Describe that scene in the picture. What equipment, including clothes, would you need for a day's sailing? Would you like to take the place of one of the men? Give your reasons.

Reasons for the popularity of sailing. Increasing use of the sea for leisure activities. Need for provision of more Marinas. Competitive racing versus family pastime.

(from: Archer, M. and Nolan-Woods, E. 1977:21)

TEST ITEM SIX

Listen to the following short dialogues. Then look at the suggestions below each as to what kind of situation they occur in. Choose the one you think best and put a cross through its number.

- a) Voice A: It's not that time, is it?
Voice B: Yes, it is. 7:30. Must get up.
Situations: 1. bed time
2. early morning
3. supper time
4. closing time
- b) Voice A: Can I speak to Mr. Brown?
Voice B: Yes, I'll put you through.
Situations: 1. a party
2. an introduction
3. outside a house
4. on the telephone

(from: Ingram, E. 1977:96)

APPENDIX 2

GLOSSARY

ACHIEVEMENT TEST – a test that measures the extent to which a person has achieved something, acquired certain information, or mastered certain skills – usually as a result of planned instruction or training. (Collins, 1969:192)

APTITUDE TEST – a measure of the student's probable performance in learning a foreign language, showing whether the student has any special aptitude for learning a new language. (Heaton, 1979:XX)

ATTAINMENT TEST – (see **ACHIEVEMENT TEST**)

BACKWASH – effects of a test on teaching. If a test has good backwash effects, it will exert a good influence on the learning and teaching that takes place before the test. (Heaton, 1979: X)

CLOZE TEST or **CLOZE PROCEDURE** – a test-construction procedure that involves deleting words on some systematic basis and replacing the deletion with blanks which the learner must fill in. (Oller, 1980:30)

COMMUNICATIVE COMPETENCE – the social knowledge and skill which permits the learner to produce and comprehend socially appropriate utterances. (Bell, 1981:151)

COMMUNICATIVE PERFORMANCE – the actual use of the language which is formally

possible, implementationally feasible, contextually appropriate and actually occurring. (Based on Hymes, 1979:5 - 26)

COMMUNICATIVE TEST – one that is based not only on a selection of items chosen on linguistic grounds alone, but on the testee's needs and requirements to use the language. The criterion for such an effective test lies not on formal correctness, but on communicative effectiveness. (Based on Carroll, 1980: 78)

COMPLETION ITEM – a test item calling for completion (filling in) of a phrase, sentence, etc. from which one or more parts have been omitted. (Collins, 1969: 207)

CONCURRENT VALIDITY (see also **EMPIRICAL VALIDITY**) – the extent to which scores on the test are in agreement with some given criterion measure. (Collins, 1969: 205)

CONSTRUCT VALIDITY (see also **VALIDITY**) – the quality of a test that is capable of measuring certain specific characteristics in accordance with a theory of language behaviour and learning. This type of validity assumes the existence of certain learning theories or constructs underlying the acquisition of the abilities or skills. (Heaton, 1979: 154)

CONTENT VALIDITY (see also **VALIDITY**) – the extent to which the content of the test represents a balanced and adequate sampling of the outcomes (knowledge, skills) of the course or instructional program it is intended to cover. (Collins, 1969: 205)

CORRECTION FOR GUESSING – a reduction in score for wrong answer, sometimes applied in scoring true-false or multiple-choice questions. (Collins, 1969: 193)

CORRELATION – relation or going-togetherness between two sets of scores of measures; tendency of one score to vary concomitantly with the other, as the tendency of students of high IQ to be above average in reading ability. The existence of strong relationship, i.e., a high correlation between two variables does not necessarily indicate that one has any causal influence on the other. (Collins, 1969: 193)

CRITERION REFERENCED TEST – a test that describes the behaviour of an individual with reference to externally predetermined and specified objectives. (Imgram, 1977:26)

CURSORY READING – a general term to denote skills involved in reading quickly, skimming and scanning. (Heaton, 1979: 124)

DEVIATION – the amount by which a score differs from some reference value, such as the mean, the norm, or the score on some other test. (Collins, 1969: 194)

DIAGNOSTIC TEST – a test primarily designed to assess the student's knowledge and skills in particular areas before a course of study is begun. (Heaton, 1979: XI)

DISCRETE POINT TEST – an analytic language test based on the notion that one and only one point of grammar should be tested at a time, that one and only one skill should be assessed at a time, and that productive and receptive repertoires, as well as oral and visual repertoires, should be tested separately. (Oller, 1980: 30)

DISCRIMINATING POWER – the ability of a test item to differentiate between persons possessing much or little of some trait. (Collins, 1969: 194)

DISCRIMINATION – (see **DISCRIMINATING POWER**)

DISTRACTOR – any incorrect choice (option) in a test item. (Collins, 1969:195)

DISTRIBUTION (frequency distribution) – a tabulation of the scores (or other attributes) of a group of individuals to show the number (frequency) of each score, or of those within the range of each interval. (Collins, 1969:195)

EMPIRICAL VALIDITY or **STATISTICAL VALIDITY** – obtained as a result of comparing the results of the test with the results of some criterion measure such as: (i) an existing test, known or believed to be valid and given at the same time; or (ii) the teacher's ratings or any other form of independent assessment given at the same time; or (iii) the subsequent performance of the testees on a certain task measured by some valid test; or (iv) the teacher's ratings or any other such form of independent assessment given later

(i) & (ii) = concurrent validity

(iii) & (iv) = predictive validity (Heaton, 1979:154)

ENABLING SKILLS – lower order intrasentential and segmental skills (e.g. lexical or phonetic phenomena) developed in order to master the language mechanisms needed for the exercise of communication skills which have to be mobilized to complete a global communicative task (e.g. search texts for specific information). They may be identified by an analysis of performance in operational terms. (Based on Carroll, 1980:33)

EXPECTANCY GRAMMAR – the learner's predictive competence in formal, functional and strategic terms. (Based on Oller, 1979:XVI)

EXTRAPOLATION – any process of estimating values of a variable data. (Remmers, 1966:373)

FACE VALIDITY – the way the test looks to the examinees, test administrators, educators and the like. (Harris, 1969: 21) Even if a test possesses all the other characteristics but lacks face validity, it will not be selected and its manifest good points will never be known. (Bell, 1981: 199)

FREQUENCY DISTRIBUTION – the number of times each of the possible events occurs in each of the possible categories. (Harris & Hodges, 1981:124)

GUESSING – responding to a question in a random manner on a choice-type test or answering a question when not absolutely certain of the answer. (Collins, 1969: 208)

INTEGRATIVE TEST – a test which attempts to test several elements and perhaps skills at the same time. (Bell, 1981:195)

ITEM – a single question or exercise in a test. (Collins, 1969:198)

ITEM ANALYSIS – the process of evaluating single test items in respect to certain characteristics. It usually involves determining the difficulty value and the discriminating power of the item, and often its correlation with some external criterion. (Collins, 1969: 198)

MATCHING ITEM – an item calling for the correct association of each entry in one list with an entry in a second list. (Collins, 1969: 209)

NORM-REFERENCED TEST – a test that compares the level of performance of an individual

with the general standard of performance which is shown by the total group that he belongs to and can be compared with. (Ingram, 1977:26)

NORMAL DISTRIBUTION – a distribution of scores or measures that in graphic form has a distinctive bell-shaped appearance (normal probability or Gaussian curve). In such a normal distribution, scores or measures are distributed symmetrically about the mean, with as many cases up to various distances above the mean as down to equal distances below. (Collins, 1969:199)

OBJECTIVE TEST – one that is scored rather mechanically without need to evaluate complex performance on a scale. (Lado, 1964:28)

PERFORMANCE TEST – any test that calls for responses that are an actual work sample of the activity being measured. (Harris & Hodges, 1981:233)

PLACEMENT TEST – a test used to assign persons to different levels of instruction, as a college English placement test. (Harris & Hodges, 1981:241)

POSTTEST – the assessment of learning at the end of an experiment, a learning task, or an instructional period. (Harris & Hodges, 1981:245)

PRAGMATIC TEST – any procedure or task that causes the learner to process sequences of elements in a language, that conform to the normal contextual constraints of that language, and which requires the learner to relate sequences of linguistic elements via pragmatic mappings to extralinguistic contexts. (Oller, 1979:38)

PREDICTIVE VALIDITY (see also **VALIDITY**) – the extent to which scores on the test predict some given criterion measure. It refers to the accuracy with which an aptitude, prognostic or readiness test indicates future learning success in some area, as evidence by correlations between scores on the test and future criterion measures of such success. (Collins, 1969:205)

PRETEST – a test given before instruction or experiment. Specifically: a. a test comparable to a posttest to be administered later; b. a test designed to help interpret instruction or experimental behaviour, as a test of ability. (Harris & Hodges, 1981:248)

PROFICIENCY TEST – one that looks back over previous language learning, the precise details of which are probably unknown, with a view to possible success in some future activity, not necessarily language learning but requiring the effective use of language. (Broughton et alii, 1980:158)

PROGNOSTIC TEST – (see **APTITUDE TEST**)

PROGRESS TEST – the assessment of the progress which students have made in mastering the material taught in the classroom. (Heaton, 1979:X)

RELIABILITY – the extent to which a test is consistent with itself in measuring whatever it does measure. (Remmers, 1966:337)

RELIABILITY COEFFICIENT – the coefficient of correlation obtained between two forms of a test (alternate-form or parallel-form reliability); between scores on repeated administration

of the same test (test-retest reliability); between halves of a test properly corrected (split-half reliability); or by using the Kuder-Richardson formula (Remmers; 1966:377)

SAMPLING – the act or process of selecting a limited number of observations, individuals or cases to represent a particular universe. (Good, 1959:475)

SCANNING – the skills used when reading to locate specific information. (Heaton, 1979:124)

SKIMMING – to read rapidly and selectively, but purposefully, rather than to read carefully. (Harris & Hodges, 1981: 298)

STANDARD DEVIATION – (see **DEVIATION**)

STATISTICAL VALIDITY – (see **EMPIRICAL VALIDITY**)

STEM – the initial part of each multiple-choice item. (Heaton, 1979:14)

SUBJECTIVE TEST – one that requires an opinion, a judgement on the part of the examiner. (Lado, 1964:28)

TRUE-FALSE ITEM – an objective test item with two response options only: T or F. (Harris & Hodges, 1981:337)

USAGE – is concerned primarily with formal language patterns. (Carroll, 1980:7)

The citation of words and sentences as manifestations of the language system. (Widdowson, 1980:18)

The aspect of performance which makes evident the extent to which the language user demonstrates his knowledge of linguistic rules. (Widdowson, 1978:3)

USE – is concerned with communicative function, with how language is used. (Carroll, 1980:7)

It is the way the system of the language is realized for normal communication purposes. (Widdowson, 1978:18)

It is the aspect of performance which makes evident the extent to which the language user demonstrates his ability to use his knowledge of linguistic rules for effective communication. (Widdowson, 1978:3)

VALIDITY (see also **CONTENT VALIDITY**, **CONCURRENT VALIDITY**, **PREDICTIVE VALIDITY**, **FACE VALIDITY**, **EMPIRICAL VALIDITY**) – the extent to which a test does the job for which it is used. This definition is more satisfactory than the traditional “extent to which a test measures what it is supposed to measure”, since the validity of a test is always specific to the purposes for which the test is used. The term validity then, has different connotations for various types of tests and thus, a different kind of validity evidence is appropriate for each. (Collins, 1969: 205)

WASHBACK – (see **BACKWASH**)

APPENDIX 3

ANNOTATED BIBLIOGRAPHY

The following annotated bibliography does not claim to be comprehensive; it is just an indication of books and a few articles that the authors have read during study seminars on testing.

Instead of simply giving the bibliographical references, it was decided to mention the contents of each book. In this way this brief list may serve as a first step to those who intend to go into a deeper study in the field.

1. ALLEN, J.P.B. & DAVIES, A. (1977) (ed.) *Testing and experimental methods*. Edinburgh Course in Applied Linguistics, v. 4. London, OUP. 233 pages.

This book represents a selection of statistical procedures that may be used by students of Applied Linguistics. In order to show the relation between testing and experiments the book is divided into six chapters: 1. Introduction; 2. Basic concepts in testing; 3. The construction of language tests; 4. The design and interpretation of experiments; 5. Procedures and computations in the analysis of experiments; 6. Statistical inference.

The contributors are: Alan Davies, Elisabeth Ingram, and Ruth Clark.

2. BELL, R.T. (1981) *An introduction to applied linguistics: approaches and methods in language teaching*. London, Batsford Academic and Educational. 271 pages.

The whole book deals with some ideas in Applied Linguistics which influence the teaching and learning of English as a foreign language. It has two main parts: 1. Syllabus design – the key issues and 2. A century of controversy in language teaching.

There are four appendices: a. Job analysis and ESP; b. Contrastive analysis and error analysis; c. Language teaching materials. Appendix C expands the general principles of test construction and use in the evaluation of courses and individuals. It also gives examples of language tests.

3. BROUGHTON, G. *et alli*. (1980) Assessment and examination. In: BROUGHTON. *Teaching English as a foreign language*. 2 ed. London, Routledge & Kegan Paul, p. 145-65.

It is a chapter that explains in a simple way some of the main points a language teacher should know about assessment and examination. Its organization is objective and provides an overview of definitions and examples of the different aspects of assessment (test). The subdivisions of the chapter are: Basic terms; Subjective and objective testing; Discrete item tests; Dictation; Cloze tests; Testing communication; Four kinds of assessment; Test qualities; Public examinations; Suggestions for further reading.

4. CARROLL, B.J. (1980) *Testing communicative performance: an interim study*. Oxford, Pergamon Press. 149 pages.

A book aiming at showing the influence of recent language teaching methodology on testing. The author sees a need to find ways of testing communicative teaching (language seen essentially as a tool for communication). The book is divided into five chapters: 1. The assessment of communicative performance; 2. The design of communicative tests; 3. The development of

communicative tests; 4. Operating the test instrument; 5. Communicative testing literature. The main ideas of the book have been developed according to experience the author gained on projects in the British Council. On reading the book one notices the influence of works by John Munby and Roges Hawkey, also from the British Council.

5. COLLINS, H.W. (1976) *Educational measurement and evaluation: a worktext*. 2.ed. Glenview, Scott, Foresman and Company. 285 pages.

It is an introduction to the techniques of measurement and evaluation. Each chapter has a list of reference and at the end there is a glossary of measurement terms. There are nine chapters: 1. Instructional objectives; 2. Planning, constructing and administering teacher-made tests; 3. Types of test items; 4. Statistical treatment of test scores; 5. Using and improving teacher-made tests; 6. Observation and student involvement in the evaluation process; 7. Standardized measurement; 8. The measurement of noncognitive learning; 9. Grading and reporting. Appendices: a. Glossary of measurement terms; b. Formulas; c. Areas of normal curve; d. Extraction of square roots; e. Table of square roots; f. Selected test publishers.

6. CORDER, S. P. (1973) *Introducing applied linguistics*. Aylesbury, Hazell Watson & Viney. 392 pages.

The author applies general accepted principles from Linguistics to language teaching. The first two parts of the book present aspects of Linguistics considered as prerequisites for the study of Applied Linguistics. Part three restricts Applied Linguistics to language teaching studies. Its contents are:

Part 1 – Language and language learning: 1. Views of language; 2. Functions of language; 3. The variability of language; 4. Language as a symbolic system.

Part 2 – Linguistics and language teaching: 5. Linguistics and language teaching; 6. Psycholinguistics and language teaching; 7. Applied linguistics and language teaching; 8. The description of languages.

Part 3 – The techniques of applied linguistics: 9. Selection 1: Comparison of varieties; 10. Selection 2: Contrastive linguistic studies; 11. Selection 3: The study of learner's language – error analysis; 12. Organization: The structure of the syllabus; 13. Presentation: Pedagogic grammars; 14. Evaluation, validation and tests.

7. DAVIES, A. (1968) (ed.) *Language testing symposium: a psycholinguistic approach*. London, OUP. 214 pages.

There are four main sections in the book: 1. The basic disciplines and their relevance to language testing, e. g. evaluation, linguistics and psychology (chapters 2, 3, 4); 2. Uses and types of tests, e. g. achievement, aptitude, oral tests and mother-tongue tests (chapters 5, 6, 7, 8); 3. The influence of tests on education (chapters 9, 10, 11); 4. The item analysis needed (appendix).

The contributors are: Keith Brown, J.B. Carroll, Alan Davies, D.W. Grieve, D.P. Harris, Elisabeth Ingram, G.E. Perren, A. E.G. Pilliner, Paul Pimsleur and Andrew Wilkinson.

8. GROUNDLUND, N. (1968) (ed.) *Readings in measurement and evaluation*. London, Collier-Macmillan. 455 pages.

It is a selection of articles ranging from basic elements of the measurement process to current issues. It has eight parts. Each one begins with a general overview of the selection content, followed by at least five articles on the corresponding topic. Parts: 1. The measurement and

evaluation process; 2. Constructing classroom tests; 3. Interpreting test scores and norms; 4. Validity and reliability; 5. Selecting standardized tests; 6. Standardized testing; 7. Using the result of measurement; 8. Trends, new developments and current issues.

Some of the contributors are: Henry S. Dyer, Paul L. Dressel, David Kratwohl, Lee J. Cronbach, J. Thomas Hasting, Robert Ebel, Leo Nedelsky, Dana Kurfman, Sherman Tinkelman, Warren Findley, Robert Thorndike, Martin Katz, David Goslin, John Flanagan.

9. HARRIS, D.P. (1969) *Testing English as a second language*. New York, McGraw-Hill. 149 pages.

A practical book for the ESL teacher to improve classroom measures and to make assessment of standardized tests with discrete items. The book covers a wide range of topics, from the general principles to the administering and evaluation of a test. The chapters are: 1. Purposes and methods of language testing; characteristics of a good test; 3. Testing grammatical structure; 4. Testing auditory discrimination and comprehension; 5. Testing vocabulary; 6. Testing reading comprehension; 7. Testing writing; 8. Testing oral production; 9. Constructing the test; 10. Administering the test; 11. Interpreting and using test results; 12. Computing some basic test statistics.

10. HEATON, J.B. (1975) *Writing English language tests: a practical guide for teachers of English as a second or foreign language*. London, Longman. 236 pages.

A book to help teachers on the construction of their English tests. In addition to outlining the general principles of language testing, the book shows the teacher how to construct a wide range of discrete point tests. Several types of test items are described. It is divided into ten chapters: 1. Introduction to language testing; 2. Objective testing; 3. Tests of grammar and usage; 4. Testing vocabulary; 5. Listening comprehension tests; 6. Oral production tests; 7. Testing reading comprehension; 8. Testing the writing skills; 9. Criteria and types of tests; 10. Interpreting test scores.

11. LADO, R. (1964) *Language testing: the construction and use of foreign language tests*. New York, McGraw-Hill. 389 pages.

This is the main testing exemplar of the structuralist view of language and language learning. It is a comprehensive book which explains and exemplifies analytical, discrete point tests. Lado's book is followed by a series of books, also mentioned in this bibliography, such as Davies (1968), Harris (1969), Heaton (1975) and Valette (1967). The book is divided into five parts:

Part 1 – General introduction and theory of foreign language testing: 1. Language; 2. Language learning; 3. Language testing; 4. Variables and strategies of language testing; 5. Critical evaluation of tests.

Part 2 – Testing and elements of language: 6. Pronunciation: the sound segments; 7. Testing recognition of the sound segments; 8. Testing production of the sound segments. 9. The testing of "stress"; 10. The testing of intonation; 11. Testing control of the grammatical structure of a foreign language; 12. Testing production of the grammatical structure of a foreign language; 13. Testing vocabulary.

Part 3 – Testing the integrated skills: 14. Auditory comprehension; 15. Reading comprehension in a foreign language; 16. Speaking a foreign language; 17. Writing a foreign language; 18. The testing of translation; 19. Testing over-all control of the language.

Part 4 – Beyond language: 20. How to test cross-cultural understanding; 21. Testing the higher values.

Part 5 – Refining and using foreign language tests: 22. Norms; 23. Validity; 24. Reliability; 25. Item analysis, equivalent forms, scoring; 26. Achievement, diagnostic, and aptitude testing; 27. Designing experiments in foreign language learning.

12. MORROW, K. (1979) Communicative language testing: revolution or evolution. In: BRUMFIT, C.J. & JOHNSON, K. *The communicative approach to language teaching*. London, OUP, p. 143-57.

It is a polemical paper that deals with the problem of measuring communicative proficiency. The author believes that communicative proficiency can be measured and suggests that earlier testing procedures need to be reviewed and reconsidered.

13. OLLER, J.W.Jr. (1979) *Language testing at school: a pragmatic approach*. London, Longman. 492 pages.

The whole book tries to answer two main questions: how can language testing relate to a pragmatic view of language as communication and how can language testing relate to educational measurement in general? Oller questions the utility of discrete point testing and recommends pragmatically-oriented language tests. He emphasizes the need of effective testing procedures which reflect a communicative view of language learning and teaching, but which can also be within the teacher's control. The introduction in chapter one gives a general view of the subject. Part one defines the requirements on pragmatic testing. Part two defines and criticizes discrete point tests and part three exemplifies and justifies alternative pragmatic tests. Each chapter is followed by a list of key points and discussion questions plus suggested readings.

1. Introduction.

Part one – Theory and research bases for pragmatic language testing.

2. Language skill as a pragmatic expectancy grammar;
3. Discrete point, integrative or pragmatic tests;
4. Multilingual assessment;
5. Measuring attitudes and motivations.

Part two – Theories and methods of discrete point testing.

6. Syntactic linguistics as a source for discrete point methods;
7. Statistical traps;
8. Discrete point tests;
9. Multiple choice tests.

Part three – Practical recommendations for language testing.

10. Dictation and closely related auditory tasks;
11. Tests of productive oral communication;
12. Varieties of cloze procedures;
13. Essays and related writing tasks;
14. Inventing new tests in relation to a coherent curriculum. Appendix: The factorial structure of language proficiency: divisible or not?

14. _____ & RICHARDS, Jack C. (1973) (ed.) *Focus on the learner: pragmatic perspectives for the language teacher*. Rowley, Newbury House. 306 pages.

A book of readings on practical theories of language and learning. It is divided into six parts: 1. The relevance of linguistics and psychology to language teaching; 2. Language learning processes; 3. Aspects of second language learning; 4. Aspects of testing; 5. Sociocultural and motivational factors; 6. Alternatives to formal language instructions.

Contributors: William F. Mackey, H. H. Stern, N. Chomsky, John W. Oller Jr., John Macnamara, Graeme D. Kennedy, H. V. George, Bernard Spolsky, John Upshur, Eugene J. Brière, R. C. Gardner, W. E. Lambert, G. R. Turner, E. F. O'Doherty, Gerald Dykstia, Sliho S. Nunes, Thomas M. Hale, Eva C. Budar.

15. OLLER, J.W.Jr. *et alii.* (1978) *Language in education: testing the tests.* Rowley, Newbury House. 135 pages.

The book investigates language proficiency as a factor in educational tests. It is recommended to course topics such as educational measurement, the psychology of language, educational linguistics, and language testing. The chapters are: 1. How important is language proficiency to IQ and other educational tests? 2. A look at the content similarities between intelligence achievement, personality, and language tests; 3. Cloze and dictation tasks as predictions of intelligence and achievement scores; 4. Relationship among oral and written cloze scores and achievement test scores in a bilingual setting; 5. Language proficiency as a source of variance in self-reported affective variables.

Contributors: John Oller, B. Gunnarson, Thomas A. Stump, Virginia Streiff, Kyle Perkins.

16. REMMERS, H. (1965) *A practical introduction to measurement and evaluation.* 2 ed. New York, Harper & Row. 385 pages.

It is a basic orientation for beginners in field of educational testing. Its contents are:

Part one – Orientation: 1. Purpose and organization of this book; 2. Survey of evaluation in the school program; 3. Statistical concepts used in measurement.

Part two – The school testing program: 4. Development and administration of the evaluation program; 5. Selection of measurement instruments; 6. Administration, scoring, and interpretation of tests.

Part three – Evaluation of classroom instruction: 7. Identification of educational objectives; 8. Constructing teacher-made tests; 9. Assigning marks and reporting pupil progress.

Part four – Appraisal of personality aspects: 10. Determining attitudes and interests; 11. Assessing emotional and social adjustment. Appendix A: list of test publishers. Appendix B: Glossary of common measurement terms.

17. VALETTE, R.M. (1967) *Modern language testing: a handbook.* 1. ed. New York, Hartcourt. 200 pages.

It shows a structuralist view of language testing used in the 1960s. Its aim is to introduce teachers to a diversity of testing techniques based on the teaching and testing theories of the time. The book has two main parts:

Part one – Principles and procedures: 1. Testing - its role in the classroom; 2. Preparing the test; 3. Giving the test; 4. Evaluating classroom test results.

Part two – Methods of evaluation: 5. The listening test; 6. The speaking test; 7. The reading test; 8. The writing test; 9. Culture and literature.

