

# 英语测试实用教程

杨端和 于燕京 编著  
刘 钦 审校

## A PRACTICAL COURSE IN ENGLISH TESTING

云南大学出版社  
Yunnan University Press  
1992

滇新登字 07 号

责任编辑 李继毛  
封面设计 杜 伟

英语测试实用教程

杨端和 于燕京 编著

云南大学出版社发行  
(昆明南泰学校内)

昆明市中印刷厂印刷

开本 187×109 1/32 印张 10.9 字数 245 千

1992 年 4 月第 1 版 1992 年 4 月第一次印刷

印数: 0001-2700

ISBN7-81025-189-9 / H·5 定价: 5.40 元

## Foreword

As the Southwest China English Teacher Training Center, housed in Yunnan Normal University and financially supported by UNESCO, has begun to enroll teacher-trainees from Sichuan, Guizhou and Yunnan Provinces and Guangxi Zhuang Autonomous Region, we feel an urgent need to offer an English testing course at the center. In our opinion, testing is a weak link in our country's foreign language teaching and learning. Many language teachers do not have any knowledge about testing and evaluation. Every school term they write various test papers of their own, but they do not know if their tests are good or not, nor can they use statistical means to analyze the test results. In offering this course, we have assumed the task of writing a course book for our teacher trainees at the center.

In this course book, we intend to outline the general principles of language testing; to classify English tests according to different criteria and purposes; to exemplify a wide range of testing techniques; to demonstrate how these techniques can be effectively used to assess the four macro-skills, namely, listening, speaking, reading and writing; to verify how various statistical methods may be applied in evaluating test results; and to help our teacher-trainees

improve their own test-writing skills.

Needless to say, in compiling a course book of this kind, we have drawn heavily upon some other works in the same line during the writing, and these works have been listed in the Main Bibliography at the end of the book as our acknowledgements.

We wish to express our gratitude to Professor Gui Shi-chun of Guangzhou Institute of Foreign Languages, without his enlightening course in Educational Design, Testing and Statistics one of the compilers followed when he sought for an M.A. in the Institute, this book would have never been written. Our indebtedness is due to Dr. D.V.Ingram, Chief Examiner of Australia. When we attended a training course in Griffith University, Australia, he personally gave us a series of lectures on language testing and allowed us to have access to those papers on testing written by him and many books he possessed. We are grateful to Professor Yang Chuan-zhe, Dean of Foreign Languages Department, Yunnan Normal University, who has showed great concern over the compiling of this book and always offered invaluable suggestions and comments. We also want to thank Mr. Li Ji-mao, the young and promising editor at Yunnan University Press, who is directly responsible for the publication of this book and whose persistent effort has made this book published ahead of schedule.

It was not until we had begun the writing that we suddenly realized that our effort was actually the first attempt to

write an English testing course book in our country. Since such is the case, and since this book has been written in a hurry to meet an urgent need, mistakes and improper arguments may have crept in at one or other places. We sincerely hope testing experts, our colleagues and students will make comments and suggestions on how to improve this book.

Finally, if our students should consider this book useful and take it as a "lead-in" to their further study of English testing, we would feel our efforts amply rewarded.

Yunnan Normal University

Kunming

April 1992

The compilers

# Contents

## Chapter 1 English Tsting, Some Basic Concepts

|                                                    |    |
|----------------------------------------------------|----|
| 1.1 Introduction .....                             | 1  |
| 1.2 types of tests classified .....                | 3  |
| 1.3 Characteristics of a good test .....           | 10 |
| 1.4 The principal English testing techniques ..... | 21 |
| Exercises and Review Questions .....               | 37 |

## Chapter 2 statistics in English Testing

|                                               |    |
|-----------------------------------------------|----|
| 2.1 Introduction .....                        | 41 |
| 2.2 Frequency distribution .....              | 41 |
| Exercises for Section 2.2 .....               | 47 |
| 2.3 Mean, median and mode .....               | 48 |
| 2.4 The variance and standard deviation ..... | 50 |
| 2.5 Item analysis .....                       | 52 |
| 2.6 The correlation coefficient .....         | 58 |
| Exercises for Section 2.5 and 2.6 .....       | 64 |
| 2.7 Calculating reliability .....             | 65 |
| 2.8 Guessing .....                            | 68 |
| Exercises for Section 2.7 and 2.8 .....       | 70 |
| 2.9 The normal distribution .....             | 71 |
| Exercises for Section 2.9 .....               | 76 |
| 2.10 Statistical significance testing .....   | 77 |
| Exercises for 2.10.4 .....                    | 84 |

### Chapter 3 Testing Listening Comprehension

|                                                      |     |
|------------------------------------------------------|-----|
| 3.1. Some problems .....                             | 85  |
| 3.2. Specifying listening tasks .....                | 86  |
| 3.3. Phoneme identification and discrimination ..... | 87  |
| 3.4. Testing suprasegmental phonemes .....           | 91  |
| 3.5. Using pictures and maps .....                   | 96  |
| 3.6. Testing the ability to guess .....              | 104 |
| 3.7. Using multiple-choice technique .....           | 106 |
| 3.8. Testing note-taking ability .....               | 112 |
| Exercises and discussion questions .....             | 117 |

### Chapter 4 Testing Spoken English

|                                              |     |
|----------------------------------------------|-----|
| 4.1. Components in spoken language .....     | 131 |
| 4.2. Paper-and-pencil tests .....            | 133 |
| 4.3. Reading aloud .....                     | 136 |
| 4.4. Scored interviews .....                 | 139 |
| 4.5. Picture use .....                       | 147 |
| 4.6. The short talk .....                    | 150 |
| 4.7. Group discussion and role playing ..... | 152 |
| Exercises .....                              | 153 |

### Chapter 5 Vocabulary Testing

|                                                 |     |
|-------------------------------------------------|-----|
| 5.1. The importance of vocabulary testing ..... | 157 |
| 5.2. Testing active vocabulary .....            | 157 |
| 5.3. Testing passive vocabulary .....           | 165 |
| 5.4. Selection of the test words .....          | 181 |
| Exercises and discussion questions .....        | 185 |

### Chapter 6 Testing Grammatical Structure

|                                          |     |
|------------------------------------------|-----|
| 6.1. Determination of test content ..... | 188 |
|------------------------------------------|-----|

|                                                        |     |
|--------------------------------------------------------|-----|
| 6.2. Mistake identification items .....                | 189 |
| 6.3. Word-order items .....                            | 192 |
| 6.4. Completion items .....                            | 193 |
| 6.5. Transformation items .....                        | 197 |
| 6.6. Multiple-choice grammatical structure items ..... | 199 |
| Exercises and discussion questions .....               | 212 |

### Chapter 7 Testing Reading Comprehension

|                                                   |     |
|---------------------------------------------------|-----|
| 7.1. What is meant by reading comprehension ..... | 215 |
| 7.2. Sentence identification .....                | 218 |
| 7.3. True / false reading tests .....             | 220 |
| 7.4. Completion items .....                       | 226 |
| 7.5. Cloze technique .....                        | 230 |
| 7.6. Multiple-choice items .....                  | 234 |
| 7.7. Testing fast reading .....                   | 253 |
| 7.8. Assembling the final form .....              | 260 |
| Exercises and Questions .....                     | 261 |

### Chapter 8 Testing Writing Ability

|                                                                |     |
|----------------------------------------------------------------|-----|
| 8.1. What is meant by "writing" .....                          | 265 |
| 8.2. Composition and objective writing tests .....             | 266 |
| 8.3. Some objective writing tests .....                        | 269 |
| 8.4. Paragraph writing .....                                   | 277 |
| 8.5. Letter writing .....                                      | 279 |
| 8.6. Extract adaptation .....                                  | 282 |
| 8.7. Writing tests based on pictures, tables or diagrams ..... | 286 |
| 8.8. Preparation of the writing test .....                     | 293 |
| 8.9. Scoring the writing test .....                            | 295 |

|                                                                          |     |
|--------------------------------------------------------------------------|-----|
| Exercises and discussion questions .....                                 | 297 |
| Chapter 9 Constructing and Administering the test                        |     |
| 9.1 Planning the test .....                                              | 306 |
| 9.2 Preparing the test items and directions .....                        | 314 |
| 9.3 Reviewing the items .....                                            | 317 |
| 9.4 pretesting the material .....                                        | 317 |
| 9.5 Analyzing the pretest results (item analysis) .....                  | 319 |
| 9.6 Assembling the final form .....                                      | 321 |
| 9.7 Reproducing the test .....                                           | 322 |
| 9.8 Using separate answer sheets for multiple-choice<br>item tests ..... | 324 |
| 9.9 Preparing equivalent forms .....                                     | 327 |
| 9.10 Preparing for the test .....                                        | 328 |
| 9.11 Conducting the testing preliminaries .....                          | 332 |
| 9.12 Conducting the test .....                                           | 335 |
| Appendix                                                                 |     |
| Table A — Area Under the Normal curve .....                              | 338 |
| Table B — “t” -distribution .....                                        | 339 |
| Main Bibliography .....                                                  | 340 |

# Chapter 1 English Testing, Some Basic Concepts

## 1.1 Introduction

Recent developments in Communicative language teaching are now resulting in a widespread reappraisal of language tests and techniques. Not only have the shortcomings of the traditional approach to language teaching and testing been strongly attacked, but the whole psychometric basis of language testing has been seriously questioned. However, we should take care not to discard every well-tried and proven method of testing in our search for some magic formula or new technique which will enable us to solve our problems in the assessment of language used as communication. All too often in language testing, as in language teaching, there seems to be a tendency for many to accept half-formed theories readily and apply them hastily and uncritically. It is always important for our English teacher-trainees to keep the best of established methods while, at the same time, seeking to develop where necessary new and more appropriate techniques to reflect the different emphases now being placed on language learning.



It might be emphasised at this stage that a concentration solely on such aspects of communicative competence as authenticity, appropriacy and register is inappropriate if the testing of grammatical system of the language is neglected and regarded as subordinate or inferior in any way. Language still consists of grammar. While testing grammatical competence within a model of language learning, however, we should pay more attention to ways in which the selection and production of a language form is determined not only by its grammatical correctness but also by its function within a given communicative area.

While it is always important to experiment with new testing techniques, it is also essential to attempt to develop existing techniques much further. After falling into disrepute in the 1960's, the long-established dictation test has been the subject of considerable research during the last decade. At the same time, a number of criteria for evaluating spoken language have been suggested, and multiple-choice items and cloze procedures have been introduced as new techniques for assessing performance on advanced reading tests. However, as the American linguist J.W. Oller points out: "There is no basis to conclude that a single test such as a dictation or a cloze test is the best way.....A multiplicity of testing methods concentrating on the kinds of language tasks which language users will be expected to perform makes the best language test in any given set of circumstances."

## 1.2 Types of tests classified

Language tests can be classified according to various criteria. There is some confusion regarding the terminology used to denote the different types of language tests in use, but a number of test specialists agree on the following broad division.

### 1.2.1 Tests classified according to the purpose

Types of language tests can be classified according to various purposes for which the tests are written. These purpose-oriented tests are again subdivided into the following four kinds:

#### 1) *Achievement tests*

These tests can be further subdivided into class progress tests and (standardised) achievement tests.

*Class progress tests* Most teachers are, at some time or other, required to construct various class progress tests. Each progress test situation is unique and can only be evaluated fully by the class teacher in the light of his knowledge of the students, the program which they have been following, and his own particular aims and goals. It is illogical to expect general purpose tests and books of tests to function as effectively as a test constructed specially for a particular situation: therefore, one of the purposes of this textbook is to help the English teachers construct their own language tests.

The progress test is designed to measure the extent to

which the students have mastered the material taught in the classroom. It is based on the language programme which the class has been following and is just as important as an assessment of the teacher's own work as the student's own learning. Results obtained from the progress test enable the teacher to become more familiar with the work of each of his students and with the progress of the class in general. The class progress test is a teaching device, its backwash effect on teaching and motivation being important features. A good progress test should encourage the student to perform well in the target language in a positive manner and to gain additional confidence. Its aim is to stimulate learning and to reinforce what has been taught, therefore, the progress test is chiefly concerned with allowing the student to show what he has mastered. Scores on it should thus be high (provided, of course, that progress has indeed been made).

*Achievement tests* Achievement (or attainment) tests, though similar in a number of ways to progress tests, are far more formal tests and are intended to measure achievement on a large scale. Most annual school examinations take the form of achievement tests; all public tests which are intended to show mastery of a particular syllabus are also achievement tests. These tests are based on what the student is presumed to have learned—not necessarily on what he has actually learned nor on what has actually been taught. Achievement tests frequently take the form of secondary school tests and school certificate examinations; many are

based on a published syllabus and exert a strong influence on its effectiveness in schools. Constructors of such tests rarely teach any of the students being tested (often an advantage provided that the test constructors are very familiar with the teaching and learning problems of the testees). Indeed, this is often a prerequisite before anyone can be appointed to any position of responsibility in connection with this type of test, though this principle obviously cannot always be applied to school examinations.

Several achievement tests are standardised: they are pretested, each item is analysed and revised where necessary, norms are established and comparisons made between performances of different students and different schools. Since such tests are administered year after year, it is possible to compare performances of students one year with those of students taking the test another year.

## 2) *Proficiency tests*

Whereas an achievement test looks back on what should have been learned, the proficiency test looks forward, defining a student's language proficiency with reference to a particular task which he will be required to perform. Proficiency tests are in no way related to any syllabus or teaching programme; indeed, many proficiency tests are intended for students from several different schools, countries and even different language backgrounds. The proficiency test is concerned simply with measuring the student's control of the language in the light of what he will be expected to do with it

in his future performance of a particular task. Does the student know enough English, for example, to follow a certain university or college course given in the medium of English? Does he know enough English in order to function efficiently in a particular type of employment for which he has applied? The proficiency test is thus concerned with measuring not general attainment but specific skills in the light of the language demands made later on the student by his future course of study or job. For instance, English Proficiency Test (E P T) constructed in our country, Test Of English as a Foreign Language (TOEFL) in the United States, the Cambridge English Certificate Test in Great Britain and Australia Second Language Proficiency Ratings (ASLPR) in Australia are all English proficiency tests to assess if a candidate knows enough English to survive and to follow courses in English-speaking countries.

### 3) *Aptitude tests*

A language aptitude test is designed to measure the student's probable performance in a foreign language which he has not started to learn: i.e. it assesses his aptitude for learning a language. Language learning aptitude is a complex matter, consisting of such factors as intelligence, age, motivation, memory, phonological sensitivity and sensitivity to grammatical patterning. The relative weighting given to these elements must depend on many factors and thus vary considerably from one individual to another. Some specialists in this field maintain that it is neither possible nor desir-

able to take an overall measurement of language aptitude; consequently aptitude is sometimes divided into various aspects according to the specific tasks for which a person is being trained: e.g. listening, interpreting, translating. Aptitude tests generally seek to predict the student's probable strengths and weaknesses in learning a foreign language by measuring his performance in an artificial language. The ability to learn new phonemic distinctions and also to use language patterns in an unfamiliar but logical manner is tested by means of the artificial language.

### 4) *Diagnostic tests*

Although the term diagnostic test is widely used, few tests are constructed solely as diagnostic tests. Achievement tests and proficiency tests, however, are frequently used for diagnostic purposes: areas of difficulty are diagnosed in such tests so that appropriate remedial action can be taken later. Sections of tests which lend themselves particularly well to diagnostic purposes are phoneme discrimination tests, grammar and usage tests, and certain controlled writing tests. Clearly, weaknesses indicated in a test of vocabulary are not highly significant in themselves and can only be regarded as indicating general weaknesses. Similarly, many existing tests of reading comprehension are not very suitable for diagnostic purposes. Tests of writing and oral production can be used diagnostically provided that there is an appreciation of the limits to which such tests can be put. Since diagnosing strengths and weaknesses is such an important fea-

ture of progress tests and of teaching, the teacher should always be alert to every facet of achievement revealed in a class progress test.

### 1.2.2 Tests classified according to the evaluation of marks

Tests can also be classified on the basis of evaluating marks. Under this category, we can divide tests into two major kinds.

#### 1) *Norm-referenced tests*

Norm-referenced tests take a certain 'norm' to evaluate a candidate's marks. The norm is, in fact, demonstrated by the mean and the standard deviation in a set of scores obtained by a group of homogeneous candidates. Therefore, norm-referenced tests are meant to evaluate a candidate's scores by referring to other candidates' scores on the same tests. These tests are designed and used to pick out good candidates among a large group of people. Without comparing one person's scores with those of the others, it is impossible for us to see who are better than others or who is the best in the group. For example, if a student scored 75 marks in an achievement test, but he ranks the fourth in the class of 25. In this way, we can conclude he is one of the upper-level students in the class.

#### 2) *Criterion-referenced tests*

Instead of referring to a norm (the mean of a group of scores), criterion-referenced tests have a pre-determined criterion or goal, so it is also called goal-referenced tests. The

criterion in these tests is absolute. It will not change even if the majority of candidates score lower than it. In other words, if all the candidates score higher than the criterion, they all should be considered to have passed the test. For example, TOFEL is often used by some American universities as a criterion-referenced test. The University of Michigan Program in Linguistics stipulates that students whose native language is not English must submit scores on the TOEFL. Scores of at least 620 are necessary to be admitted into the Program. Here the scores 620 is the criterion predetermined by the authorities of the Program.

### 1.2.3 Tests classified according to methods of scoring

If we classify English tests on the basis of scoring methods then we can divide different tests into the following two kinds.

#### 1) *Subjective tests*

The name subjective tests implies that in scoring such tests, the scorer's subjective judgement plays an important part in scoring the test results. Question-answering, comment-making and composition-writing are the most commonly used subjective tests. Such tests have a long history in language teaching, so they are often called traditional tests.

#### 2) *Objective tests*

In scoring such a test, no any scorer's subjective factors are involved. No matter who the scorer is, the marks for the same test should be the same. Objective tests can be scored

by an optical scanner connected to a computer. In this way, so much time and so many material resources can be saved. For this reason, in large-scale tests, objective test items form a greater part. In fact, in the last twenty years, many testing theories and methods have been developed to write and evaluate objective items. Multiple choice question and true-false judgement are the two most frequently used objective test means.

Both subjective and objective tests have their own merits and demerits. Without a context, we cannot say one is better than the other.

### 1.3 Characteristics of a good test

All good tests possess three qualities: validity, reliability and practicality. That is to say, any test that we use must be appropriate in terms of our objectives, dependable in the evidence it provides, and applicable to our particular situation. To be sure, there are other test characteristics which are also of value, but these three constitute the most essential ones, without any one of which a test would be a poor investment in time and money. Whether the teacher is constructing his own test or is selecting a standard instrument for use in his class or school, he should certainly understand what these concepts mean and how to apply them.

Although most treatments of the subject put validity first, as being the most important of the three elements, we

shall begin with reliability for the reason that it generally affects validity and that validity cannot be fully appreciated without a basic understanding of reliability.

#### *Reliability*

By reliability is meant the stability of test scores. A test cannot measure anything well unless it measures consistently. To have confidence in a measuring instrument, we would need to be assured, for example, that approximately the same results would be obtained (1) if we tested a group on Tuesday instead of Monday; (2) if we gave two parallel forms of the test to the same group on Monday and on Tuesday; (3) if we scored a particular test on Tuesday instead of Monday; (4) if two or more competent scorers scored the test independently. It is clear from the foregoing that two somewhat different types of consistency or reliability are involved: reliability of the test itself, and reliability of the scoring of the test.

Reliability is of primary importance in the use of both public achievement and proficiency tests and classroom tests. Methods of estimating the reliability in a test will be indicated in Section 2.7 of chapter 2. However, an appreciation of the various factors affecting reliability is important for the teacher at the very outset, since many teachers tend to regard tests as infallible measuring instruments and fail to realise that even the best test is indeed an imprecise instrument with which to measure language skills.

Factors affecting the reliability of a test are:

- (1) the extent of the sample of material selected for testing: whereas validity is concerned chiefly with the content of the sample, reliability is concerned with the size. The Larger the sample (i.e. the more tasks the testee has to perform), the greater the probability that the test as a whole is reliable, hence the favouring of objective tests by some people, which allow for a wide field to be covered.
- (2) the administration of the test: is the same test administered to different groups under different conditions or at different times? Clearly, this is an important factor in deciding reliability, especially in tests of oral production and auditory comprehension.
- (3) test instructions: are the various tasks expected from the testee made clear to all candidates?
- (4) personal factors such as motivation and illness.
- (5) scoring the test: one of the most important factors affecting reliability. Objective tests overcome this problem of marker reliability, but subjective tests are still faced with it: hence the importance of the work carried out in the field of the multiple-marking of compositions.

In fact, our short discussion above concerning reliability was intended simply to explain what is meant by the term and to suggest the importance of high reliability in the preparation or adoption of an educational measure. Whenever one is considering the adoption of a standard test, he should take into account the estimates of reliability which the publisher provides, and the nature of the groups on which these

estimates are based. Likewise, when locally produced tests are to be used repeatedly to make important decisions about candidates, it is essential to obtain an estimate of the reliability of the instrument. On the other hand, for the usual classroom test, prepared by the class teacher and used but once, the computing of reliability coefficients is scarcely ever practicable. What is important for the classroom teacher, however, is to understand the chief factors affecting reliability and why and how some kinds of testing techniques are inherently more reliable than others. In our chapters on the testing of the various language skills, we shall make frequent references to the problems of obtaining satisfactory test and scorer reliability.

Finally, it must always be remembered that reliability refers purely and simply to the precision with which the test measures. No matter how high the reliability quotient, it is by no means a guarantee that the test measures what the test user wants to measure. Data concerning what the test measures must be sought from some source outside the statistics of the test itself. This problem will be considered in the following section.

### *Validity*

In the selection of any test, two questions must always be considered: (1) What precisely does the test measure? and (2) how well the test measure? If the test is found to be based upon a sound analysis of the skill or skills we wish to measure, and if there is sufficient evidence that test scores corre-

late fairly highly with actual ability in the skills area being tested, then we may feel reasonably safe in assuming that the test is valid for our purposes. A number of types of validation are applied to tests, all of them attempting to answer the above questions. In this brief survey we shall limit our consideration to just a few of the most common kinds.

*Face validity* The following test item is invalid if we wish solely to measure writing ability: "Is photography an art or a science? Discuss." It is likely to be invalid simply because it demands some knowledge of photography and will consequently favour certain students. In fact, even a superficial inspection of the above essay item would be sufficient to reveal that it was not valid. This type of validity is, therefore, often referred to as face validity: if a test item looks right to other testers, teachers, moderators, and testees, it can be described as having at least face validity. Hence it is often useful to show a test to colleagues and friends. As constructors of the test, we can become so involved in the test that we sometimes fail to stand back and look at the individual test items objectively. Only if the test is examined by other people can some of the absurdities and ambiguities then be discovered.

Language tests which have been designed primarily for one country and are adopted by another country may lack face validity. A vocabulary or reading comprehension test containing such words as 'typhoon', 'sampan', 'abacus', and 'chopsticks' will obviously not be valid in Eastern Africa no matter how valid and useful a test it has proved in China.

The same argument applies to many standardised tests designed for immigrants in the United States and used now in many parts of the world. Although no substitute for empirical validity, face validity can provide not only a quick, reasonable guide but also a balance to too great a concern with statistical analysis. Motivation is maintained if a test has good face validity, for most testees will try harder if the test looks sound.

*Content validity* This kind of validity depends on a careful analysis of the language being tested and of the particular course objectives. The test should be so constructed as to contain a representative sample of the course, the relationship between the test items and the course objectives always being apparent. There is a strong tendency, especially in multiple-choice testing, to test only those areas of the language which lend themselves readily to testing. Many tests of phonology, for instance, concentrate too much on testing phoneme discrimination rather than the more important features of stress and intonation: one cannot help but suspect the reason for this is simply that phoneme tests are often easier to construct than items testing stress and intonation.

*Construct validity* If a test has construct validity, it 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 abilities and skills. For example, it can be argued that a speed reading test

based on a short comprehension passage is an inadequate measure of reading ability (and thus has low construct validity) unless it is believed that the speed reading of short passages relates closely to the ability to read a book quickly and efficiently and is a proven factor in reading ability. If the assumption is held that systematic language habits are best acquired by means of the structural grammar approach, then a test which emphasises the lexical or situational meaning of language (rather than the structural meaning) will have low construct validity.

*Empirical validity* A fourth type of validity is usually referred to as statistical or empirical validity. This validity is obtained as a result of comparing the results of the test with the results of some criterion measure such as:

- (1) an existing test, known or believed to be valid and given at the same time; or
- (2) the teacher's ratings or any other such form of independent assessment given at the same time; or
- (3) the subsequent performance of the testees on a certain task measured by some valid test; or
- (4) the teacher's ratings or any other such form of independent assessment given later.

The test situation or the technique used is always an important factor in determining the overall validity of any test. Although an ideal test situation will by no means guarantee validity, a poor test situation will certainly detract from it. Is an auditory comprehension test valid if the testee hears only

a disembodied voice on, say, a poor quality tape-recorder?

Finally, it should be noted that a test can be reliable without necessarily possessing validity. However, reliability is clearly inadequate by itself if a test does not succeed in measuring what it supposed to measure.

### *Practicality*

A third characteristic of a good test is its practicality or usability. A test may be highly reliable and valid but still be beyond our means of facilities. Thus in the preparation of a new test or the adoption of an existing one, we must keep in mind a number of very practical considerations.

*Economy* As most educational administrators are very well aware, testing can be expensive. If a standard test is used, we must take into account the cost per copy, and whether or not the test books are reusable. Again, it should be determined whether several administrators and / or scorers will be needed, for the more personnel who must be involved in scoring a test, the more costly the process becomes.

Closely related to economy in money is economy in time. In writing or selecting a test, we should certainly pay some attention to how long the administering and scoring of it will take. Yet in choosing a short test rather than a longer one, we must proceed with caution. As we observed earlier in this chapter, both reliability and validity depend in a large part on the adequacy of our sampling and the breadth of our appraisal. Our task, then, is to select an instrument which is of sufficient length to yield dependable and meaningful re-



sults but which will also fit comfortably into the time that can be made available for testing, this point is of particular importance when the test must be administered in the classroom and scored by the classroom teacher.

*Ease of administration and scoring* A test must be fairly easy to administer if it is practical. It is only too easy for testers to become so absorbed in the actual construction of test items that the most obvious practical considerations concerning the test are overlooked. The length of time available for the administration of the test is frequently misjudged even by experienced test-writers, especially if the complete test consists of a number of sub-tests. In such cases sufficient time may not be allowed for the administration of the test, the collection of the answer sheets, the reading of the test instructions, etc. In the case of all large-scale tests, the time to be allowed should be decided on as a result of a pilot administration of the test (i.e. a try-out of the test to a small but representative group of testees).

Another practical consideration concerns the answer sheets and the stationery used. Many tests require the testee to enter his answers on the actual question paper (e.g. circling the letter of the correct option), thereby unfortunately reducing the speed of the scoring and preventing the question paper from being used a second time. In some tests, the candidate is presented with a separate answer sheet, but too often insufficient thought has been given to possible errors arising from the (mental) transfer of the answer from the

context of the item on the question paper to the answer sheet itself. Confusion may result, for example, if the items are numbered vertically on the question papers and horizontal numbering is adopted for the corresponding answer sheet.

E.G.:

1. You'd already left by seven o'clock,.....you?  
A. didn't b. weren't c. hadn't d. haven't
2. If you take swimming lessons, you.....soon.  
A. will be able to swim C. can swim  
B. swim D. shall have swum
3. Did anyone tell Tom.....the careless error he made?  
A. off B. over C. on D. about

Circle the letter of the correct answer.

1. A B C D 2. A B C D 3. A B C D
4. A B C D 5. A B C D 6. A B C D

The use of separate answer sheets, however, greatly facilitates marking and is strongly recommended when large numbers of students are being tested.

It is of paramount importance that examiners are fully conversant with the test situation. If the test is to be administered by several examiners working in different test centres, clear directions specifying exactly what each examiner should say and do should be issued in order to ensure that exactly the same procedure is followed in each centre. Although this principle seems obvious enough, it is extremely difficult in many cases for the test writer to see his test from the point of view of the person conducting the test, simply because he is