REVIEW OF RELEVANT LITERATURE
CHAPTER 2

REVIEW OF RELEVANT LITERATURE

Testing has its roots in psychology from where it has come to education in general and language testing in particular. In its earlier manifestation, testing was used in education and psychology for the purpose of measuring a person's knowledge, intelligence, aptitude, attitude, motivation, other characteristics etc in a systematic way. The use of testing became more popular than it was in the times after World War I and World War II, particularly so after the Government of the USA started organising specialised language training programme for its armed forces. The tradition methods used in language testing that consisted largely of essay-type test items gave way very gradually to objective-type tests. Language tests are used today for the purpose of diagnosing language deficiency in learners, or to measure their present level of language proficiency, and so on. This has led to the availability of many types of language tests. However, researchers have taken note of what is called the 'g' factor.

We consider the testing of academic achievement important because we need more tests of this kind that have been authenticated through empirical research. The testing of achievement of a testee on any
given syllabus or course unit is important because all other decisions with regard to his/her continuance in higher education and suitability for employment are dependent on the outcome of such tests. Established theorists on the subject like Baker (1989), Lado (1961, 64), Valette (1977), Henning (1987), and Hughes (1989), to name but a few, apparently find achievement tests to be unworthy of their attention. They give this a small amount of printing space, and even that, not separately, but as a part of their discussion on the types of language tests, or discuss them as a part of some other related chapter. We need to understand what this test called achievement test really is. A given syllabus (also referred to as Course Content in recent times) gives us a clear idea of (a) the aim, (b) the objectives, (c) the content, (d) the methodology, (e) the materials - primary and secondary, and (f) the mode of evaluation, to be used on a given course. What is tested is the attainment of the aim and the objectives set for the course to achieve. An achievement test seeks "to measure the extent of learning in a prescribed content domain" and this is done with reference to the "explicitly stated objectives" of a learning programme and the results could be used for evaluating the course and/or the "certification of learned competence" (Henning 1987: 6). We have yet to see a better definition of achievement tests and would, therefore, like to use it for the purpose of this research work. We would like to mention an important work in this area.
and that is one by Canale (1985) who has focused focuses on achievement testing, albeit with a tilt toward proficiency-orientation. This research focuses its attention on the testing of learners’ achievement in English and, as said earlier, there does not seem to be adequate research work available in this area with regard to the Indian context, except Khan (1987) and Raviya (2002). However, even these works are but surveys conducted using a checklist-cum-questionnaire. The present research work not only uses a survey method but also a testing package designed to measure achievement including related skill areas like study skills. It is but natural that something of this kind would bring with it its own share of constraints as well as freedom for this researcher. Having said that, we would like to make a quick review of the literature available to us in language testing.

Khan (1987), Raviya (2002), and this researcher undertook research in this vital area within ELT. Khan, who worked on a minor research project for the UGC on the state of language testing in Compulsory English at the tertiary level in Gujarat, and Raviya, who subsequently worked on his doctoral work under him as a part of the team with this researcher, give a good account of the state of language testing in Gujarat in their unpublished work.
In the concluding chapter of his doctoral work, Raviya, concurring with the findings of Khan (1987), observes that the following four null hypotheses they had set up for the research stand accepted:

1. The testees' familiarity with the test format and the structure of the test-tasks in vogue for long allows them to easily guess the questions likely to figure in any examination by going through the question papers in that examination in the last four or five years.

2. The structure of the test-tasks does not allow the tester to measure the linguistics, communicative, organisational, textual, and strategic competencies of the testees, along with their ability to argue cogently, and cohere their ideas into a unifying whole using cohesive devices.

3. The testees generally ignore the textbooks prescribed for them and used for the purpose of setting question papers in any examination, with the result that they fail to provide convincing evidence of their familiarity with the texts.

4. The excessive reliance on comparative marking adds to lack of reliability of the tests. The term comparative marking refers to
reconsideration of a number of scripts attracting similar marks and slight revision of the marks awarded in some cases.

We have tried our best to cover it all here as per the following outline: (1) Primary Books and Handbooks, (2) Concepts in Language Testing (Measurement, Validation, Reliability, Variability, Methodology, Techniques, Statistics, Item Response Theory (IRT), Needs Analysis, Curriculum, English for Specific Purposes, and Translation), (3) Test and Test Construction, (4) Test Types (Proficiency Tests, Communicative Language Testing, Oral Interview, Cloze Test), and (5) Study Skills.

Primary Books and Handbooks

Cronbach published an important book on psychological testing in 1949 in the aftermath of World War II, which set a flurry of activity. The emergence of the US as a superpower created great interest in it in the psychology of learning and the learning of foreign languages. This interest subsequently accelerated the pace of research in language testing. The development of the Taxonomy of Educational Objectives (Bloom, Englehart, Furst, Hill, and Kratwohl 1956) was a giant step in this direction. Harris (1969) followed the lead given by Cronbach (1949) by publishing *Testing English as a Second Language*. 

56

Carroll and Hall (1985) seem to have followed the lead given by Heaton (1976, 1982) in publishing a Handbook of language testing called *Make Your Own Language Tests: A Practical Guide to Writing Language Performance Tests*. The following year the American Council on

There has been a flurry of activity in the world of research on language testing. Language testing continues to focus predominantly on objective type of test items rather than essay type test-tasks, with the sole exception of the Test of Written English (TWE) from Educational Testing Services in the USA and the Writing Module, which is a part of Academic

**Concepts in Language Testing**

There are numerous concepts that can be discussed under this broad heading. Some of these are Measurement, Validation, Reliability,
Review of Relevant Literature

Variability, Methodology, Techniques, Statistics, Item Analysis, Item Response Theory (IRT), Needs Analysis, Curriculum, English for Specific Purposes, and Translation. We shall begin the discussion in this section with the idea of measurement.

Measurement

Measurement is an important concept in language testing. Bachman (1991) devotes an entire chapter in his book *Fundamental Considerations in Language Testing* to discuss the concept of *measurement*, which involves the quantifying achievement, progress, proficiency etc in terms of grades and percentages. Popham (1978) focuses attention on criterion-referenced measurement, whereas Briere and Hiindofotis (1979) examines concepts in language testing with reference to some recent studies. If, on the one hand, Canale and Swain (1980) broke fresh ground by deliberating on the theoretical bases of communicative approaches to second language teaching and testing, Bachman and Palmer (1981) undertake a multitrait multimethod investigation into the construct validity of six tests of speaking and reading on the other. It is in a context of this kind that Wesche (1981) seeks to take a good look at communicative testing in a second language, and Bachman (1982) explored the subject of the trait structure of cloze test scores.
Other studies like Lewkowicz (1983) takes a hard look at method effect in testing reading comprehension by making a comparison of three methods. Tarone (1983) does some loud thinking on the notion of ‘communication strategy’. Traub (1983) undertakes a study titled “a priori considerations in choosing an item response model”. Even so, there was a constant shift back to basics as in case of Bachman and Palmer (1984) who zero in on comments the terminology of language testing by offering their valued comments. It is indeed heartening to know that statistical concepts like item analysis and test scoring with binary logistic models (Mislevy and Bock 1984) are also given due research attention.

Shohamy, who has carried out considerable research in language testing begins by addressing the question of whether or not the testing method makes a difference in reading comprehension reading comprehension (Shohamy 1984). She goes on to collaborate with Reeves to ask questions like ‘where from’ and ‘where to’ about authentic language tests (Shohamy and Reves 1985). Spolsky (1985) lends his might by investigating into the limits of authenticity in language testing, and so do Porter, Floden, Freeman, Schmidt, and Schwille (1986) by exploring what they call “content determinants”. Thus, the 1980s were productive insofar as language testing is concerned.
Spolsky (1986) gives us food for thought when he engages the teachers among us in dealing with a multiple choice for language testers, and Rafaldini (1988) grapples with the question of the use of situation tests as measures of communicative ability. However, there appear to be regular intervals when researchers like, Linn (1989), for example, return to the concept of educational measurement. Wiggins (1989) mulls over the idea of moving toward more authentic and equitable assessment, which, as the researcher says, is what a true test is all about. Happily, there is research on communication strategies and their significance in foreign language teaching (Willems 1987), creativity and the possibility of simultaneous interpretation (Alexieva 1990), principles of test theories (Suen 1990), fundamental considerations in language testing (Bachman 1991), strategic competence and how to teach it (Dornyei and Thurrell 1991) etc but a return yet again to educational measurement as in case of the sampling variability of performance assessments (Shavelson, Baxter, and Gao 1993). Shohamy calls our attention to the question of the power of tests, especially with regard to the impact of language tests on teaching and learning (Shohamy 1993), and Wall and Alderson (1993) examine the washback based on the Sri Lankan Impact Study.
The latter part of the 1990s is apparently devoted to a variety of areas in research like a study of washback effect vis-à-vis TOEFL preparation courses (Alderson and Hamp-Lyons 1996), learning styles and strategies (Ely and Pease-Alvarez 1996), tasks for independent language learning (Gardner and Miller 1996), the reconceiving of test theory (Mislevy 1996), testing and experimental methods (Allen and Davies 1997), making classrooms more responsive to diversity through peer-assisted learning strategies (Fuchs, Fuchs, Mathes, and Simmons 1997), ethical concerns about washback, impact and validity (Hamp-Lyons 1997), impact and washback in language testing (Wall 1998), and another look at some basic language testing concepts (Thrasher 2000).

Validation

The concern with validation of language tests has been of serious concern for researchers. For example, Campbell and Fiske (1959) focus attention on the convergent and discriminant validation by the multitrait multimethod matrix, and Cronbach (1971) with test validation, Cronbach (1980) with how one could go straight if validity were to be on parole, Bachman and Palmer (1981) with the construct validation of the FSI oral interview, and Clifford (1981) with the need for a research model vis-à-vis convergent and discriminant validation of integrated and unitary
language skills. This preoccupation with validation is in evidence even in
the 1980s when Palmer, Groot and Trosper (1981) examined the idea of the
constructed validation of tests of communicative competence, Bachman
and Palmer (1982) joined hands to research the construct validation of
some components of communicative proficiency, but it was not until 1984
that the validity of using holistic scoring to evaluate writing gets under the
scanner (Charney 1984). Davies (1984) goes into the process of validation
of three tests of English language proficiency only to be confronted by Low
(1984) looking into the validity and the problem of direct language
proficiency, and Stevenson (1985) probing pop validity and performance
testing.

A context of this kind is obviously responsible for Groyjahn
(1986) investigating into some methodological considerations involving the
use of cognitive psychology to look at the process of test validation. An
important landmark in this regard is the final report on ELTS (now IELTS)
Validation Project by Criper and Davies (1987). If, on the one hand, De
Jong and Glas (1987) subsequently examine the validation of listening
comprehension tests using item response theory, then Bachman (1988)
focus on problems in Examining the validity of the ACTFL Oral
Proficiency Interview, Messick (1989a, 1989b) on the science and ethics of

**Reliability**

The study of validity and validation is inseparably linked with reliability of our tests in English. Therefore, reliability is one of the areas that have occupied attention. We shall present a sample of studies on this aspect of language testing. Cronbach (1947) investigated into the meaning and determination of test reliability. McBride and Martin (1983) directed attention on reliability and validity of adaptive ability tests in a military setting. Thissen (1990) draws our attention to reliability and measurement precision. Wainer et al (1990) offer to us a primer on computerized adaptive testing, Jamieson et al (1993) concentrate on the reliability of a computerized scoring routine for an open-ended task, and Conlan et al (1994) make a study of intra-rater reliability of assessments of live versus audio-recorded interviews in the IELTS speaking component for the Inter-
national Editing Committee of IELTS. However, Merrylees and McDowell (1999) break new ground by making an investigation into the reliability of speaking test with particular reference to the attitude of the examiner to the format of the speaking test and the discourse produced by a candidate/examiner. Another area of interest in language testing is the concept of variability.

Variability

Variability has been studied variously. For instance, Ellis (1987) studies the style shifting in the use of the past tense in the context of interlanguage variability in narrative discourse, Long (1983) introduces us to LISREL by taking a look at the covariance structure models, Oller and Perkins (1978) examine intelligence and language proficiency as sources of variance in self-reported affective variables, and Skehan (1987) reviews the idea of variability and language testing. In retrospect, there have been studies such as the one by Spurling and Ilyin (1985), which takes a hard look at the impact of learner variables on language test performance, Tarone (1983, 1988) reports on the variability in interlanguage systems, and variation in interlanguage, and Wolfram (1985) makes a case for the obvious on variability in tense marking. Even so, there is a marked tendency to explore the methodology of testing.
Review of Relevant Literature

Methodology

Review of Relevant Literature

Techniques


Statistics


Alongside these, there are others like Mansfield (1986), who discusses with us basic statistics and its applications. Mislevy and Bock (1986) examine the concept of item analysis and test scoring using binary logistic models. Stansfield (1986) attempts a study on the face validity of the multitrait multimethod matrix and the convergent and discriminant validity of oral proficiency tests. Woods, Fletcher, and Hughes (1986) unfold for us the use of statistics in language studies. Powers (1986) is concerned with the academic demands related to listening skills. Reid (1986) examines the use of the writer's workbench in composition teaching and testing. Pollitt and Hutchinson (1987) uncover the process of calibrating graded assessment using Rasch partial credit analysis of performance in writing. Wainer, and Kiely (1987) makes a case for testlets for item clusters and computerized adaptive testing. We also have Andrich (1988) detailing for us Rasch models for measurement at a time when

testing, and Statistics Corner (2001) discusses questions and answers about statistics in language testing. This brings us to the concept of item response theory (IRT).

IRT

Item Response Theory is an important area within statistics in language testing. Hambleton (1983) focuses on the application of item response theory. Hambleton and Swaminathan (1985) discuss the principles and applications of item response theory. An important study in this area is by Henning, Hudson, and Turner (1985) who examine item response theory and the assumption of unidimensionality for language tests. Woods and Baker (1985) and Hambleton, Swaminathan, and Rogers (1991) help us revisit the concept of item response theory. Having done this, we will need to survey the developments in needs analysis and curriculum studies.

Needs Analysis

A syllabus must answer a need or a set of needs if it were to be considered to be need-based. That is why needs analysis is a significant concept in syllabus and curriculum design today. A 'need' is something that is felt and something the fulfilment of which one cannot do without. Our learners require to master basic competencies like language skills, subject-
specific skills, study skills, referencing skills, note-taking and note-making skills, time management skills etc. That is why Walter (1983) offers us a rich fare of the processes involved in identifying the language needs of overseas students in tertiary education in the United Kingdom in an unpublished doctoral work submitted to the University of London, and Weir (1983) apparently works alongside on the same subject for his doctoral degree at the University of London.

Curriculum


English for Specific Purposes

Work on English for Specific Purposes can be traced back to as late as 1965 when Herbert (1965) presents us with the structure of technical English. Glendinning (1975) follows with focus on the structure...
of English in Mechanical Engineering. However, it was in the 1980s that Mackay and Palmer (1981) reveal the basic of program design and evaluation in languages for specific purposes. Closer home, Singh (1982) unveils a plan for the teaching of English in technical institutions. Beretta (1983) makes a comparison of three tests of listening in the context of English for academic purposes at the University of Edinburgh. Associated Examining Board (1984) starts a new trend with their test in English for educational purposes. Pugh and Ulijn (1984) focus on studies and practice in native and foreign languages with reference to reading for professional purposes. It is yet again the Associated Examining Board (1987) that makes an attempt to lay down the process of test in English for educational purposes (TEEP). Also, whereas Hutchinson and Waters (1987) present the design of English for specific purposes, Alderson (1988a, 1988b) shows us how specific we can get in the testing of English for specific purposes, and testing and its administration in ESP respectively. Chamberlain and Baumgardner (1988) explore practice and evaluation of ESP in the classroom, only to be backed up by Hughes (1988) who explores the testing English for University Study. Westaway, Alderson and Clapham (1990) call our attention by talking about the directions in testing for specific purposes, while Douglas (2000) focuses on recent developments in the
assessment of language for specific purposes. All this leaves us with another major area called translation, which we propose to take up next.

**Translation**


**Test and Test Construction**

This is one area where researchers like Lado (1961) take the lead by writing on the construction and use of foreign language tests. Brooks (1967) follows the lead by enabling teachers to make their own language tests. Davies (1977) is not far behind when he writes on the
of ESL. TEEP is revisited yet again in the form of a research paper by Aldershot (1987) for the Associated Examining Board.


Despite all this work, the focus towards the end of the 1980s and the beginning of 1990s shifts back to the question of why we need better assessment (Shepard 1989). Cangelosi (1990) and Gronlund (1993)
offer us a good fare in talking about the need for designing tests for evaluating student achievement, and on how to make achievement tests and assessments. Weir (1993) returns once again to discuss why we need to understand and develop language tests. Also, there are guidelines on how to prepare for the test of English as a foreign language (Sharp 1994). Alderson, Clapham, and Wall (1995) take us back to consider language test construction and evaluation. The Clapham study into the effect of background knowledge on reading comprehension with regard to the development of IELTS is an important addition to the literature on the subject for its sheer practical demonstration (Clapham 1996).

Kitao and Kitao (1996) enable us to consider the testing of writing, and this work is accessible on the Internet on the website of the Internet TESL Journal as per the details in the Bibliography. Cumming (1997) also addresses the question of testing writing in a second language. Gronlund (1998) examines the process of constructing achievement tests. Kitao and Kitao (1999) show us how to write a good test. Thrasher (2000a, 2000b) enables us to take a good look at test theory and test design, and test design and construction respectively. Having reviewed test construction, we shall now focus our attention on discussing test types.
Test Types

Test types are indeed an area that needs consideration. One of the first of the research works that makes an appearance goes back to the end of 1960s. Sako (1969) deals with the writing of proficiency and achievement tests. Bormuth (1970) expends a considerable amount of time in examining the theory of achievement test items. Oller (1971) explores dictation as a device for testing foreign language proficiency but Clark (1978) is concerned with the theory and application of direct testing of speaking proficiency, a discussion that continues the following year in the form of a research piece on the writing on direct vs. semi-directs tests of speaking ability (Clark 1979). Farhady (1979) focuses attention on the disjunctive fallacy between discrete point and integrative testing.

1980s is apparently a productive year when more research work is seen in evidence. Berk (1980) presents us with a consumer's guide to criterion-referenced test reliability, and Bachman (1981) on the formative evaluation in ESP program development. The Royal Society of Arts, which is well known for RSA-CUEFL, delineates examinations in the communicative use of English as foreign language: specifications and specimen papers (Royal Society of Arts 1981). Ulibarri, Spencer, and Rivas (1981) have studied language proficiency tests and their relationship to

Hughes (1986) attempt to assay a pragmatic approach to criterion-referenced testing in foreign language. Alderson (1987) focuses on whether there exist levels of reading comprehension. McCollum (1987) reviews the idea of oral language proficiency test. Finally, Alderson (1988) zeroes in on the testing of reading comprehension skills. 1990 is known for a diverse kind of research and chief among these are Haertel (1990) on achievement tests, the technical data report on Stanford Achievement Test (1990), and primary 3, form L on Stanford Achievement Test (1992). This being what it is, the preoccupation in the 1980s is with proficiency testing.
Proficiency Tests

Upshur (1979) examines functional proficiency theory with reference to the research role it is likely to have in language tests. Cohen (1980) is concerned with assessing the use of tests for testing language ability in the classroom. Henning et al (1981) continue in this vein to focus on the assessment of language proficiency and achievement among learners of English as a foreign language. Madsen and Jones (1981) undertake an exercise to take a look at the classification of oral proficiency tests. Farhady (1982) adopts a different line is taken to study measures of language proficiency from the learner’s perspective. Ingram and Wylie (1982) are concerned with second language proficiency ratings in Australia while far away in the UK, Seaton (1982) inquires into proficiency testing for tertiary level study and training in Britain.

Of all the works we have seen to date, it is Canale (1983) postulating on some dimensions of language proficiency that opens up newer avenues for the exploration of the concept of competence that was inaugurated by Chomsky in 1965. Cummins (1983) continues in the same vein to deal with language proficiency and academic achievement. Farhady is known for his work on statistical foundations of language testing and, thus, one is not surprised to find him making public his research findings on

It is interesting to note that research on language proficiency seems to be the most striking feature of work in the 1980s. If Royal (1984) mulls over language proficiency and academic achievement, then Canale (1985) focuses on proficiency-oriented achievement testing. Chen and Henning (1985) concern themselves with linguistic and cultural bias in language proficiency tests as against Ingram (1985), who attempts an overview of some aspects of testing in assessing proficiency. Also, James (1985) is concerned with foreign language proficiency in the classroom and
beyond. Rating scales are another area of discussion in literature on language testing which is why Lowe (1985) takes what he calls 'the view from the mountain' in the context of the ILR proficiency scale as a synthesizing research principle. Competence is revisited again when Kramsch (1986) debates on the movement from language proficiency to interactional competence in communication. Lowe (1986) joins issues with Kramsch, Schulz and particularly to Bachman and Savignon to take a vast sweep in examining proficiency, panacea, framework, and process. Introducing a new comprehensive test of oral proficiency (Shohamy and Bejerano 1986) is a work that is a class in itself and so is the measurement of foreign/second language proficiency (Bachman and Clark 1987).

Byren (1987) concentrates on insights from a proficiency orientation in the context of second language acquisition. Lowe (1987) takes a look at the unassimilated history of second language proficiency assessment. Wesche (1987) is concerned with the Ontario Test of ESL as an example of second language performance testing. Lantolf and Frawley (1988) make an effort to help us understand the construct that proficiency is. Even so, Jaeger (1989) discusses the certification of student competence, and De Jong and Henning (1990) examine the test dimensionality in relation to student proficiency. The discussion here is but a sampling of
research on proficiency and this researcher makes no claim that this is exhaustive.

Communicative Language Testing

Communicative language teaching and testing gained currency largely towards the end of the 1970s but came to be accepted as a major development both in language teaching and language testing. Labov (1970) is one of those who wrote on the study of language in its social context much in the same way as Hymes who is acclaimed as a proponent of the theory of communicative competence. However, it is not until after the publication of Wilkins’s Notional Syllabuses in 1976 that research interest in communicative language teaching and testing becomes evident. Morrow (1977) discusses the techniques of evaluation for a national syllabus, Munby (1978) shakes the world with the most important of his contributions titled Communicative Syllabus Design, and Brumfit and Johnson (1979) outline the Communicative Approach to Language Teaching. The last is an edited work in which we also get to read Morrow (1979) on the idea of whether communicative language testing can be considered a revolution or evolution.

It is indeed interesting to find Douglas (1986) focusing on communicative competence and tests of oral skills, Morrow (1986) on the

*Oral Interview*

Oral Interview has been used for long in the testing of proficiency in spoken English. Hinofotis et al (1982) relates FSI oral interview scores to grammatical analysis of learner's speech. Bachman and

**Cloze Test**

One of the earliest discussions on cloze tests dates back to 1956 when Taylor (1956) elaborates on the recent developments in the use of cloze procedure. Alderson (1978) carries out a study of the cloze procedure with native and non-native speakers of English in the form of an unpublished doctoral work submitted to the University of Edinburgh, Alderson (1979) concerns himself with the cloze procedure as a measure of
proficiency in English as a foreign language, and Alderson (1980) examines the native and non-native performance on cloze tests. Kikwood et al (1980) have an important work to their credit in the form of a book titled *Matching Students and Reading Materials. A cloze Procedure Method for Assessing the Reading Ability of Students and the Readability of Textual Materials*. Brown (1983) takes a closer look at reliability of cloze tests. Levenston, Nir, and Blumkulka (1984) embark on discourse analysis in the testing of reading comprehension using cloze techniques. However, it is Brown (1993), who is left to talk about the characteristics of a natural cloze. This leads us to a discussion on a major proficiency language test, IELTS designed by the University of Cambridge Local Examinations Syndicate (UCLES).

**IELTS**

One of the important works done on International English Language Testing System (IELTS), known as English Language Testing Service (ELTS) in its first incarnation is Carroll (1978), which details the specifications for an English language testing service for and on behalf of British Council, London. The International Editing Committee of IELTS commissioned Wylie (1993) to undertake a study of the inter-rater reliability of the IELTS Speaking Test. Similarly this very same committee
commissioned Conlan, Beardsley, and Martinson (1994) to study the intrarater reliability of assessments of live versus audio-recorded interviews in the IELTS Speaking Component. Wylie and Hudson (1994) reviewed the proposed Speaking Test in the revised IELTS Test on behalf of the UCLES. Brown, and Hill (1998) publish a research study examining the interviewer style and candidate performance in the IELTS oral interview for and on behalf of UCLES. Lazaraton (1998) analyses differences in linguistic features of candidates at different levels of the IELTS Speaking Test on behalf of UCLES. Merrylees and McDowell (1999) carry out an investigation of speaking test reliability with particular reference to examiner attitude to the speaking test format and candidate/examiner discourse produced. Also, there is this IELTS annual review 1998/99 published by the University of Cambridge Local Examinations Syndicate. Lazaraton (2000) analyses the relationship between task features and candidate output for the revised IELTS Speaking Test, and IELTS (2000) discusses the revision of the IELTS Speaking Test. If IELTS is a major proficiency test in English, so is the Testing of English as a Foreign Language (TOEFL), which we shall take up next.
1982 is a very productive year insofar as the TOEFL is concerned. There are three studies discussed this year by the Educational Testing Service (ETS) (1982a, 1982b, 1982c) that offer us the oral proficiency testing manual, ETS oral proficiency testing manual, and manual for score users of the Test of Spoken English (TSE). The ETS makes available to us scoring guidelines for the Test of Written English (TWE) in 1986, and 1989. Raimies (1990) deals with the causes for concern in the use of TSE. Eignor at al (1993) initiate us to take a look at the case studies in computer adaptive test design through simulation for the ETS. Alderson and Hamp-Lyons (1996) undertake a study of washback of the TOEFL preparation courses. Eignor et al (1997) discuss the development of a scale for assessing the level of computer familiarity of TOEFL examinees for the ETS. This brings us to the end of the section on the types of tests, and we shall, therefore, now move on to examining study skills.

Study Skills

note taking from lectures. Wells (1987a, 1987b, 1987c) deals with listening and note taking; note taking from lectures, and twenty-three time-management techniques respectively. Most of the others published after the 1980s carry more or less the same information and, hence, they are not being discussed here but this brings us to the end of the chapter.

The focus of our attention in the research reported here will move away from the conventional forms of language testing and seek to understand and explain the importance of study skills in any course in English for Specific Purposes (ESP) or English for Specific Academic Purposes (ESAP) and this will, therefore, find reflection in the test package we have used.