

**AN ANALYSIS OF THE EVOLUTION OF THE IELTS TEST
AND AN INVESTIGATION OF ITS FACE VALIDITY**


by

Teresa Thiel

February 1995

Declaration


I certify that this thesis contains no material which has been accepted for the award of any other degree or diploma in any tertiary institution, and that, to the best of my knowledge and belief, it contains no material previously published or written by another person, except where due acknowledgement is made in the text of the thesis.

Signature of Author 

Teresa Thiel

Permission to Copy

I hereby give permission to the staff of the University Library and to the staff and students of the School of Education within the University of Tasmania to copy this thesis in whole or in part without reference to me. This permission covers single copies made for study purposes, subject to the normal conditions of acknowledgement.

Signature of Author 

Teresa Thiel

Abstract

First, the purpose of this study was to analyse the evolution of the International English Language Testing System (IELTS), an internationally recognized language proficiency test, within the context of developments in language proficiency testing in the last decade. Approaches to the validation of language proficiency tests are discussed with particular reference to face validity and washback validity. Second, the purpose of this study was to determine the face validity of the IELTS test from an investigation into international students' perceptions of the test. To these ends, a questionnaire was administered at the University of Tasmania to international students (n=40) who had taken an IELTS test. A focus group interview was also conducted with students who had taken both the IELTS test and the Test of English as a Foreign Language (TOEFL). It was found that the IELTS test generally had high face validity amongst the sample of international students in this study. Positive washback effects appeared to reinforce the face validity of the IELTS. Test preparation was perceived to be relevant to the target situation of academic study. It was also found that an English language proficiency score of IELTS Band 6 may be an unrealistic minimum entry level in meeting the linguistic demands of some courses of study at the University of Tasmania. Tentative conclusions from this study suggest that both the overall band score and the diagnostic profile of IELTS test results should be examined carefully when admitting international students to tertiary institutions. In addition, there are also non-linguistic factors which can affect the potential academic success of international students. Recommendations concerning the key role of English language support and study skills courses conclude this study.

Acknowledgements

I wish to acknowledge the support and guidance given
by the following people in the writing of this work:

Dr Thao Lê

Associate Professor R.J.S. Macpherson

My colleagues in the English Language Centre and International Student Office
at the University of Tasmania, Launceston

and

my dear friend and colleague

Dr Mohammed Ahmed

(International University of Japan)

Table of Contents

List of Tables	ix
Glossary of Terms Used	x
Chapter 1	Introduction
	Background to the Study 1
	Purposes of the Study 2
	Research Rationale and the Structure of this Thesis 3
	The Study in Perspective 6
Chapter 2	Analysis of an Innovation
	Introduction 8
	Background to the IELTS 8
	The Munby Model 10
	Test Specifications 12
	The Nature of Language Proficiency 13
	Test Constraints 14
	Test Content 14
	Test Methods 17
	The Application of Bachman's Framework 19
	Reporting Results 21
	An International Context 22
	Assessing an Innovation 23
	The IELTS Test 1995 24
	Concluding Note 26
Chapter 3	Literature Review
	Introduction 28
	Definitions of Language Proficiency 28

The Purposes of Language Proficiency Testing	31
English for Specific Purposes	32
Communicative Language Testing	34
A Model of Second Language Proficiency	37
Test Method Effect	39
Test Performance	40
Approaches to Validity	42
Content Validity	43
Construct Validity	44
Criterion-Related Validity	46
Predictive Validity	47
Face Validity	50
Washback Validity	52
Introspective Methods	54
Summary of Approaches to Validity	55
Norm-Referenced and Criterion-Referenced Testing	57
Summary	60
Chapter 4 Research Methodology	
Introduction	61
Background to the Investigation	61
Subjects of Research	62
Instrumentation	63
Pretesting the Questionnaire	65
Pretesting the Interview	67
Design	68
Procedure	71
Limitations of Methodology	72
Summary	74

Chapter 5	Data Analysis and Interpretation	
	Introduction	75
	Sample Characteristics	75
	Test Content	79
	Proficiency Level	83
	The IELTS Test as a Predictor	87
	Test Time	89
	Diagnostic Profile	90
	Target Situation	93
	Authenticity	94
	Subjects of Interview	94
	Target Situation of Interviewees	95
	The Nature of Language Proficiency	96
	Background Knowledge	96
	Further Comparisons	97
	Test Methods	98
	Authenticity of Tasks	100
	Final Comments	101
	Limitations of the Data Analysis and Interpretation	102
	Summary	103
Chapter 6	Tentative Conclusions and Recommendations	
	Introduction	104
	Research Findings and the IELTS Test Changes	104
	Tentative Generalizations	106
	Tentative Conclusions	109
	Recommendations to Test Users	110
	Further Research	113

Bibliography	117
Appendices	
Appendix I	130
Appendix II	131
Appendix III	132
Appendix IV	136
Appendix V	137
Appendix VI	138
Appendix VII	144
Appendix VIII	148

List of Tables and Figures

Table 2.1	The English Language Testing Service, 1980	15
Table 2.2	The International English Language Testing System, 1989	16
Table 2.3	The International English Language Testing System, 1995	24
Figure 5.1	Nationality of subjects	76
Figure 5.2	Years in Australia	77
Table 5.1	Course of study	77
Figure 5.3	Number of times IELTS test taken	78
Table 5.2	IELTS overall band scores	79
Table 5.3	Perceptions of test content	80
Figure 5.4	Modules chosen by students	82
Table 5.4	Perceptions of Band 6 by level of study	83
Table 5.5	Perceptions of test performance	85
Table 5.6	Perceptions of English language level	86
Table 5.7	Perceptions of Band 6 as realistic	88
Table 5.8	Perceptions of test time	90
Figure 5.5	Perceptions of level of difficulty of subtests	91
Table 5.9	Rank result and level of difficulty (Reading)	92
Table 5.10	Rank result and level of difficulty (Listening)	92
Figure 5.6	Importance of macroskills in target situation	93

Glossary of Terms Used

DEET	Department of Employment, Education and Training
EAP	English for Academic Purposes
EFL	English as a Foreign Language
ELICOS	English Language Intensive Courses for Overseas Students
ELTS	English Language Testing Service
ESL	English as a Second Language
ESP	English for Specific Purposes
GPA	Grade Point Average. The average score achieved by a student in examinations over a given level or levels in the course being studied.
GT	General Training module
IDP	International Development Program of Australian Universities and Colleges
IELTS	International English Language Testing System
TAFE	Colleges of Technical and Further Education
TEEP	Test in English for Educational Purposes
TOEFL	Test of English as a Foreign Language
TSE	Test of Spoken English
TWE	Test of Written English
UCLES	University of Cambridge Local Examinations Syndicate

Chapter 1

Introduction

Background to the Study

The last two decades have witnessed the development of education into an international business. Competition to attract students from overseas into English-medium educational institutions now operates on a global scale providing significant export earnings for host countries. In recent years, the flow of international students, notably from Asia, into secondary and tertiary education in English-speaking countries, including Australia, the United Kingdom, Canada and the United States, has accelerated. Social, political and economic reasons are involved. In 1993, approximately 63,000 full fee-paying international students were studying in Australia, an increase of 20% between 1992 and 1993 (DEET, 1993). 68% of students from overseas were in tertiary education and 9% were attending English Language Intensive Courses for Overseas Students (ELICOS).

Non-native speakers of English admitted to undergraduate and postgraduate courses in Australia are required to fulfil second language proficiency entry requirements, unless they attained senior high school educational qualifications in the medium of English. Therefore, significant judgements have to be made about the readiness of international students for overseas academic study. These judgements are usually based on internationally recognized standardized English language proficiency tests, primarily the International English Language Testing System (IELTS) and the Test of English as a Foreign Language (TOEFL).

The criteria for assessing second language proficiency have educational and ethical implications. Admission to the target academic environment creates high expectations

amongst international students. There are challenges in adapting to a new academic genre, together with issues of cultural adjustment and cultural differences in teaching and learning styles (Ballard and Clanchy, 1988; 1991). Moreover, linguistic difficulties experienced during overseas academic study are a major concern for international students (Ballard and Clanchy, 1988: 2; Burns, 1991; Nixon, 1993). Research has shown a frequent mismatch between the linguistic criteria for admission to tertiary institutions and the linguistic demands of academic courses (Elder, 1993; University of Adelaide, 1992: 4). Second language proficiency tests are not always reliable predictors of academic success as many variables intervene to affect the academic performance of international students in the new target situation (Gibson and Rusek, 1992: 60; Graham, 1987; Light *et al.*, 1987). Key factors affecting academic performance, apart from English language proficiency, include the educational system previously experienced, the type and duration of English language learning in the home country, the level of adjustment of a student to the new academic genre and the depth of academic background knowledge (Burns, 1991: 62). As the number of international students studying in Australia grows, tertiary institutions need to continually review the type and content of courses, teaching styles, international student support services, and the provision of English language courses available in the target situation. Taking into account these considerations, crucial decisions about the English language proficiency entry requirements for international students need to be as effective as possible.

Purposes of the Study

The research reported in this thesis has two purposes. First, it aimed to analyse the evolution of an internationally recognized second language proficiency test, the IELTS. This research was undertaken at a significant stage in the evolution of the IELTS test. The 1989 version of the test was in the process of a major revision to culminate in a new version, available in 1995, intended to assess the second language proficiency of international students.

Secondly, this study investigated international students' perceptions of the 1989 version of the IELTS test in order to determine its face validity, before the introduction of a revised and updated 1995 test version. Test takers are seldom asked by examination bodies for their perceptions of second language proficiency tests and of their test-taking experience, their views on test methods and format, and whether standardized proficiency tests appear to fairly assess their ability.

During the 1985 National Foundation for Educational Research conference, Alderson (1986) appealed for consideration of a student-centred perspective on test validity, an appeal that was reiterated at the 1990 Regional Seminar on Language Testing and Language Programme Evaluation (Alderson, 1991). He suggested that test takers should be asked to reflect on their responses to test items, and to comment on any discrepancy between their test performance and their own view, or their teachers' views, of their language ability. In other words, Alderson argued that test takers should be encouraged to contribute to future developments in language testing (1991: 21). This significant, if radical proposal, influenced the direction of the research reported here. The research aimed to consult a sample of IELTS test takers studying at an Australian university for their views on this internationally recognized second language proficiency test. Testing has major consequences for the future of individuals and yet 'very few people have actually consulted the insiders, the test takers, themselves' for their views on language testing (Alderson, 1986: 99).

Research Rationale and the Structure of this Thesis

Chapter Two is an outline of the evolution of the IELTS as an internationally recognized test. It replaced the ELTS test which had been designed to assess the English language proficiency of international students entering British universities. The analysis of the IELTS test undertaken in this thesis sought to determine the underlying construct of language proficiency

in the design and development of the test, the relationship between test content and target language use, and the communicative language learning paradigm reflected in the test. In addition, Bachman's theoretical framework (1990) of factors influencing performance on language tests is used to further the analysis of test content and test methods. Chapter Two also includes an examination of the most recent evolution in the IELTS test, a revised and updated version effective from April 1995. Major changes to the test have implications for test takers and test users and reflect a shift in the construct of language proficiency underlying the IELTS test.

Language testing evolved as an academic subfield in the 1960s due to the rapid increase in the demand for language tests to solve both practical problems and theoretical issues (Cooper, 1986: 240). Changes in test design reflect changes in theories of the nature of language and the nature of testing (Morrow, 1986: 11). Chapter Three of this thesis is a review of the relevant literature and developments in the field of language proficiency testing since the 1980s, characterised by a shift in the focus of language testing research from the 'how' of testing to the 'what' of testing (Skehan, 1988: 211). From theoretical discussions about the structure of general language proficiency, descriptive models of underlying communicative competence emerged (Canale and Swain, 1980; Canale, 1987). Bachman (1990; 1991) has extended these descriptive models into a working model of communicative language ability which includes both competence and performance. Furthermore, communicative language testing and the testing of English for Specific Purposes (ESP) have significantly influenced test design and test content in the last fifteen years. The 1989 version of the IELTS test, its forerunner the English Language Testing Service test (ELTS), and the Associated Examining Board's Test in English for Educational Purposes (TEEP) illustrate a 'needs-driven approach' to language proficiency testing (Weir, 1993: 167). Research in language testing has also drawn attention to the effect on test performance of test method (Bachman and Palmer, 1986; Shohamy, 1984) and of test content within a specialized academic context (Alderson and Urquhart, 1985a; 1985b). In addition, investigations into the effect of test

taker characteristics on language test performance (Chen and Henning, 1985; Chapelle, 1988) have revealed potential sources of test bias (Bachman, 1991: 675).

Recent research into language proficiency testing has shown a 'far greater concern ... with validity as opposed to an earlier overriding concern with test reliability' (Weir, 1993: 166). In Chapter Three, therefore, there is an overview of approaches to test validity and a discussion of the role of these approaches in the validation of the IELTS test. It is shown that the face validity of a test is often defined simply as *test appeal*. In addition, the extent to which a test appears to be relevant to the test users and test takers is dismissed as inconsequential (Cronbach, 1984, cited in Bachman, 1990: 286). Nevertheless, Bachman (1990: 288) has pointed out:

Even those who have argued against test appeal as an aspect of *validity* have at the same time recognized that test appearance has a considerable effect on the acceptability of tests to both test takers and test users.

Stevenson stated that 'face validity judgements are naïve because appearances in testing are treacherous, and well-established deceivers' (1985b: 114). However, not all experts agree. 'Face validity is becoming recognized as an increasingly important factor affecting test development and indeed test performance' (Alderson, 1986: 100). Chapter Three includes an examination of the controversial role of face validity in language proficiency testing

In responding to Alderson's appeal for a greater focus on test takers, 'their interests, perspectives and insights' (Alderson, 1991: 5), this thesis also reports on an investigation into international students' perceptions of the IELTS test. Chapter Four is a description of the methodology chosen for the study, a questionnaire survey and focus group interview undertaken at the University of Tasmania. The questionnaire survey was designed to ascertain test takers' perceptions of the appropriateness of the minimum recommended entry

requirement, IELTS Band 6, for academic study at tertiary institutions in Australia, and of the IELTS test as a predictor of academic success. The survey also attempted to determine perceptions of test content and test methods, and to establish the perceived relationship between the IELTS test and the target situation of academic study. The focus group interview was intended as a means of validating the questionnaire and of finding out test takers' comparative perceptions of two internationally recognized language proficiency tests, the IELTS and the TOEFL. An analysis of the research findings from the questionnaire survey and interview are provided in Chapter Five.

On the basis of the investigation into international students' perceptions of the 1989 version of the test, conclusions are then drawn about the face validity of the IELTS test in Chapter Six. For the purposes of this study, face validity is defined as 'the appearance or perception of the test and how this may affect test performance and test use' (Bachman, 1990: 302). Furthermore, recommendations are made for the effective interpretation of English language proficiency test scores by test users and for ways of overcoming the mismatch between minimum English language proficiency entry requirements and the linguistic demands of academic courses. Possible directions for further research are also outlined.

The Study in Perspective

There have been significant methodological advances in the tools available for analysis and validation of language tests in the last decade (Bachman, 1991). Psychometric tools, such as Item Response Theory (Henning, 1987) and criterion-referenced measurement (Hudson and Lynch, 1984), have been applied to language testing. Confirmatory factor analysis and multitrait multimethod research design have also contributed to test validation. However, qualitative methods in validating language tests have also gained increasing importance in the 1990s (Alderson 1991a; Bachman, 1990). Introspection techniques, for example, provide an

insight into the strategies used by test takers in taking tests (Cohen, 1984; Grotjahn, 1987; Nevo, 1989).

Weir (1993: 166) has pointed out three major directions for second language proficiency testing in the 1990s. Firstly, there is a need for further research to determine the washback validity of tests, that is, the 'quality of the relationship between a test and associated teaching' (Alderson and Wall, 1993: 116). Research by Hughes (1988b) showed that the 'washback' effects of second language proficiency testing can bring about innovations in classroom teaching and learning. Secondly, there is scope for further developments in the communicative or authentic approach to language testing, and in the design of tests that are performance related. Thirdly, closing the gap between language testing and language use has implications for English for Specific Purposes and English for Academic Purposes (EAP) test development which relates a specifiable context to the target situation.

In addition, test design and the context and appropriacy of test items depend to a large degree on the judgements of 'experts'. Testing is a judgemental activity (Alderson, 1991: 21; Hughes, 1986: 40). According to Alderson (1991: 22) judgements made by experts are not always accurate or reliable and, therefore, should not be assumed to reflect a consensus or the only means of establishing test validity. The test takers also need to be consulted for their views on test content and test method to establish whether tests appear to be measuring what they claim to be measuring. In Alderson's terms (1986: 98), there is a need for more consumer research in second language proficiency testing. Future innovations in language testing should perhaps aim for higher face validity amongst test takers and test users. The role of the language learner in testing can no longer be that of 'an obedient examinee, a disinterested consumer, a powerless patient or even an unwilling victim' (Canale, 1985, cited in Milanovic 1991: 136).

Chapter 2

Analysis of an Innovation

It is important that designers of tests should take account of generally accepted views of the nature of language, language proficiency, and language use (Alderson and Clapham, 1992: 149).

Introduction

In this chapter, the development of the International English Language Testing System is analysed from its inception in 1989 to a revised version of the test introduced in 1995. The necessity for an international innovation in language proficiency testing in 1989 is examined. It is suggested that the IELTS test represented an evolution rather than a revolution in second language proficiency testing. Contributory reasons were the lack of a theoretical framework for language proficiency testing at the time and test constraints imposed by the examining body. A recent evolution in the IELTS test is also analysed in this chapter. It is shown that changes in the 1995 version of the test have major implications for test takers.

Background to the IELTS

The IELTS is an English language proficiency test administered by the University of Cambridge Local Examinations Syndicate (UCLES). With increasing numbers of non-native speakers of English planning to study or train overseas in English medium institutions, the IELTS was designed to meet the need for a standardized English language proficiency test for entry to both academic courses and general training programmes.

The IELTS test, established in 1989, is jointly managed by the British Council, UCLES and the International Development Program of Australian Universities and Colleges (IDP). By 1993, there were some 200 centres in 110 countries where candidates could sit for the IELTS test. Although growing in international recognition, the IELTS does not have the widespread impact of the TOEFL test, and it mainly affects those international students planning to study in Australia and Britain. In 1993, 34,000 IELTS tests were taken worldwide, of which 15,000 were taken in Australia; this emphasizes the importance of the test for entrance to Australian academic institutions (IELTS Australia, 1994).

The IELTS test has four main aims. First, it tests the language proficiency of international students and determines whether they have the required level of competence for tertiary studies in English. Second, it aims to diagnose areas of weakness in proficiency, thus enabling appropriate courses of language instruction to be designed. Third, it is intended to provide a test of general language proficiency for students at ELICOS centres in Australia. Finally, it aims to provide an internationally recognized proficiency test which has high face validity (Ingram, 1991: 186).

As an internationally recognized test, the IELTS replaced the English Language Testing Service test. The ELTS test, introduced by the British Council and UCLES in 1980, underwent a lengthy validation study commissioned from Edinburgh University (Criper and Davies, 1988) and the results contributed to the ELTS Revision Project. This project, which was conducted jointly by Britain and Australia, included a review of the literature on testing English for Academic Purposes, a report on a needs analysis of sample target populations of international students, consultations with language testers and teachers in tertiary institutions, feedback from ELTS test centres and receiving institutions, and a conference of language researchers to debate the content and form of a new language proficiency test. A key task of the Revision Project, directed by Professor J.C. Alderson of the University of Lancaster,

was to identify and define an appropriate model of language proficiency for a revised test to replace the ELTS.

Whilst the ELTS Revision Project was in progress, a survey was conducted of twenty-two applied linguists in the UK and North America in order to reach a consensus on a model of language proficiency appropriate for a new international test. In a letter accompanying the survey, with reference to the existing ELTS test, Alderson and Clapham (1992: 167) claimed, that:

There is a feeling in the field that theories of language use and ESP have now changed, and that the test is based on outdated models. Our intention is in part to verify or question this judgement and in part to identify a more suitable theoretical model for the 1990s.

Alderson and Clapham acknowledged the disappointingly low response rate of 50% to the survey. However, their findings, outlined in this chapter, are evidence of a transition period in language testing within which the IELTS emerged in 1989.

The Munby Model

The survey of applied linguists revealed differences of opinion concerning Munby's taxonomy of enabling skills on which the test specifications for the ELTS were based (Munby, 1978). The Munby model, discussed further in Chapter Three, influenced the direction of English for Specific Purposes (ESP) in the late 1970s at a time when the communicative paradigm in second language teaching was gaining support. According to Johns and Dudley-Evans (1991: 298), the definition of ESP includes various academic Englishes and thus incorporates 'general' English for academic purposes. In their terms, ESP testing involves a distinguishable group of second language learners within 'a specific

learning context'. Munby's classification of communicative language needs for specific target situations was the basis for the design of the ELTS as an English for Academic Purposes/English for Specific Purposes test.

However, by the early 1980s, there was criticism of the Munby model as outdated and inadequate. Furthermore, Munby's taxonomy failed to provide an appropriate hierarchy of the enabling skills for test purposes. Skehan (1984: 209) was particularly critical:

The taxonomy has captured the aspects of language that seemed individually most interesting in that period, and frozen them in time as though they have status in a coherent, comprehensive model.

He further commented (Skehan, 1984: 212):

We have seen that the claims of the ELTS for construct validity are questionable, since the 'theory' on which they depend, the Munby model for syllabus design, is unsatisfactory internally, and bears insufficient relationship with reality.

In fact, a significant issue addressed by the Revision Project was the assumption that the Munby model no longer reflected current thinking in language proficiency testing. Only a minority of applied linguists in the survey supported the specification of a Munby-type taxonomy of underlying skills for the design of tasks in a new language proficiency test. Nevertheless, Alderson and Clapham (1992: 155) concluded from the ELTS Revision Project that 'there is no generally accepted, dominant paradigm or theory that can be said to have replaced the Munby model'.

Test Specifications

Test specifications should aim to state clearly a test's construct and thus, in the case of any new English language proficiency test, the underlying theory of language proficiency held by the test designers (Alderson, 1988). Test specifications should also specify target performance and provide guidelines to content and test methods for the production of parallel test versions. The 1988 validation study (Criper and Davies) had criticized the ELTS test for lack of content validity, that is, the extent to which it measured a representative sample of the language proficiency skills it claimed to be measuring. The first step in establishing content validity is to analyse test specifications. However, the Revision Project produced no clear consensus amongst researchers on a construct for test specifications. The majority of applied linguists agreed on the importance of task appropriateness in the design of a new test, and the importance of a needs analysis of a sample target population. Alderson and Clapham (1992: 163) claimed that research conducted by Weir into the academic English language needs of international students entering higher education in the United Kingdom contributed to the new IELTS test specifications. However, Weir (1990: 21) was skeptical:

The recent ... revision project has forsaken such needs analysis and instead intends to rely on post hoc 'expert' comment to judge the 'authenticity' and other aspects of the products of the test writing teams.

The IELTS project writing teams were given the task of simultaneously drafting specifications for a new test and producing sample test items. Subject specialists, teachers, applied linguists, and testers were then asked for their feedback on the operationalized specifications (Alderson, 1988). In other words, 'expert judgements' were used as a measure of test validity. The validation of the IELTS test is discussed further in Chapter Three.

The Nature of Language Proficiency

Alderson and Clapham (1992:152) cite the view of language proficiency embodied in the ELTS test as defined by the validation study report (Criper and Davies, 1988: 9-10):

Firstly, it divides proficiency in the skills dimension, having separate tests of reading, listening, writing and speaking ... Secondly, it divides proficiency into 'general' and 'study' proficiency, having a test of 'study skills' distinct from the tests of the four skills referred to above ... Thirdly, it divides proficiency on the subject dimension, providing options in the form of 'modules'.

The ELTS Revision Project survey confirmed that Oller's hypothesis of an underlying unitary competence for English language proficiency was no longer widely accepted by applied linguists. In fact, Oller (1983: 35) had already conceded that language proficiency does not simply consist of 'an exhaustive global factor'.

Although Palmer and Bachman's research (1981) into the nature of language proficiency had been published, in which they hypothesized that speaking and reading abilities are independently measurable, their research does not appear to have been influential in the design of the IELTS test. Recent studies, referred to in Chapter Three, have empirically validated the differentiation of language proficiency into skill areas at intermediate levels of second language acquisition. On the other hand, the differentiation is less clear at higher proficiency levels (Milanovic, 1991: 135). According to Alderson and Clapham (1992: 164), the new IELTS test followed a skills-based approach because:

The applied linguists seemed to be content with the concept that proficiency is divisible by skill, and there are thus tests of the four macro-skills: reading, writing, listening and speaking.

Test Constraints

Alderson and Clapham (1992: 154) claimed that the British Council set provisos which constrained the innovativeness of the IELTS. Firstly, the new test had to be administratively cheaper than the ELTS. Secondly, it had to be a shorter and more manageable test. Testing time was to be reduced from 180 to 145 minutes. Thirdly, the test was constrained by the need to administer it to large numbers of candidates in diverse international settings. Fourthly, the test had to be designed for administration and marking by clerical staff with little training. The British Council also stipulated continuity with the ELTS test for reasons of face validity.

Test Content

The ELTS test consisted of two general tests of reading skills and listening skills, and three modular academic tests of study skills, written presentation and oral interaction in five subject-specific areas, as illustrated in Figure 2.1.

In the ELTS validation study, two possible alternatives for a revised test were considered, 'one small set of general specific tests or to return to one overall proficiency test' (Davies, 1988a: 47). The former alternative was recommended. In actuality, the 1989 IELTS test maintained the division between general language proficiency and ESP, but reduced the number of subject-specific areas to three broad fields of academic study. As Ingram (1991: 186) pointed out:

While face validity would suggest that the test should have modules relevant to the specific academic or training field candidates wish to enter, the sheer diversity of those fields makes a hard version of specific purpose testing impractical.

Figure 2.1

The English Language Testing Service, 1980

(Carroll, 1983: 91)

GENERAL TESTS	ESP TESTS	TEST AREAS
<u>G1 Reading Skills</u> Multiple-choice items (40 minutes)	<u>M1 Study Skills</u> Multiple-choice items (55 minutes)	Medicine Physical Sciences Social Studies
<u>G2 Listening Skills</u> Multiple-choice items (35 minutes)	<u>M2 Written Presentation</u> Two topics (40 minutes)	Technology Life Sciences General Academic
	<u>M3 Oral Interaction</u> Structured discussion (10 minutes)	Non-academic

The 1989 version of IELTS consists of two general subtests of Listening and Speaking, academically related, and two ESP subtests of Reading and Writing in three broad fields of study referred to as academic modules A, B and C, as shown in Figure 2.2. The change from ELTS 'test areas' to IELTS 'modules' necessitated incorporating Technology under Physical Sciences, Medicine under Biological Sciences, and creating a broad Humanities and Social Sciences module. A General Training module (GT) replaced the General Academic and Non-academic test areas of the ELTS. The new IELTS test gave candidates the option of being tested in the macroskills of Reading and Writing in one specific ESP module, according to their past or future academic specialism. Alternatively, candidates could opt for a less academically demanding General Training module appropriate for vocational and non-degree training courses. The 1989 version of the IELTS thus represented a compromise between an

English language proficiency test offering an infinite number of academic modules to meet the diverse target situations of study and training, and a general EAP proficiency test, such as the TOEFL.

Figure 2.2

The International English Language Testing System, 1989

(Deakin, 1994: 2)

GENERAL TESTS	ESP TESTS	TEST MODULES
<u>Listening</u> Comprehension questions (30 minutes)	<u>Reading</u> Comprehension questions (55 minutes)	A Physical Sciences B Biological Sciences C Humanities and Social Sciences GT General Training
<u>Speaking</u> Structured oral interview (11-15 minutes)	<u>Writing</u> Task One (15 minutes) Task Two (30 minutes)	

The decision to include three broad academic modules in the IELTS test was largely due to administrative convenience rather than a reflection of research into the effect of non-linguistic background knowledge on language proficiency test performance. However, Alderson and Urquhart (1985a; 1985b) had shown quite convincingly that students perform better on tests of reading comprehension in a familiar content area, with the reservation that test method also affects test performance. Their research suggests that ESP proficiency tests are fairer,

assuming one agrees that ‘general’ reading comprehension texts will tend to favour particular groups, either culturally or in subject-related content.

Although an ESP listening subtest based on academic lectures and notetaking was considered, a general listening subtest was adopted in the 1989 version of the IELTS test. This aimed to ease the problems of administration in testing centres without language laboratories and with large numbers of candidates (Ingram, 1991: 190). In addition, Alderson and Clapham's survey of applied linguists for the ELTS Revision Project revealed no strong opinions about the inclusion of a separate grammar component in a revised language proficiency test. Even though draft tests of grammar were initially written for the IELTS, they were finally excluded as it was judged that the macroskills of reading and writing adequately cover grammatical competence.

Test Methods

The ELTS test included multiple-choice test items for reading, listening and study skills, and direct test methods for writing and speaking (see Figure 2.1). The new IELTS test reflected the trend towards communicative language testing, hence, a greater variety of testing methods; cloze, true or false, and short answers. In addition, authentic texts and tasks in a variety of genres were included to simulate the specific performance tasks required of students in tertiary level academic studies. Ingram (1991: 192) explained that ‘neutral’ rather than ‘highly discipline-specific texts’ are chosen for the Reading subtest to provide realistic contexts for a diverse target population of test takers.

A further change introduced in the 1989 version of the IELTS was the integration of reading comprehension texts and writing tasks. Test instructions encouraged candidates to evaluate the content of texts from the Reading subtest when answering task two in the Writing

subtest. Integrated skills tests have been subject to some criticism because performance on one task may depend on successful performance on another task (Weir, 1990: 85). For example, reading proficiency may interfere with a candidate's ability to complete writing tasks that build on reading comprehension texts. However, Weir (1990: 84) regarded the closer integration of listening, reading and writing tasks as more realistic of the target academic situation:

In academic life, for example, students have to read a variety of texts and process spoken discourse in lectures and seminars in preparation for writing tasks, and work is often discussed with others before a final version is presented for assessment.

Two types of questions were included in the Writing subtest; a guided task involving information transfer or reprocessing from a graph, table, or diagram, and an experience task using relevant reading comprehension passages, supported by the candidate's own experience, to present an argument or to suggest a solution to a problem. Further discussion of the development and design of the writing test component is to be found in a paper by Westaway (1988).

The General Training module for reading and writing was designed to be similar to the academic modules previously described. However, tasks are related more specifically to survival skills in an academic environment, such as writing letters, completing application forms and following instructions. In addition, reading comprehension texts are shorter, and less demanding lexically.

Whilst reading and writing are tested in an ESP context, listening and speaking are general proficiency tests representing a common core for all IELTS candidates. The Listening subtest includes the contexts of both social survival and a new study environment (see Appendix I). The use of a variety of discourse styles and a range of native speaker accents aims to reflect

authentic language use. For administrative purposes, input is aural from a pre-recorded cassette tape, and visual stimuli, such as sketch maps and drawings, may be included in the initial part of the test. Test items include note-taking, form-filling and short answers, as well as cloze and multiple-choice types. The Listening subtest does not attempt to simulate the situation of an authentic extended academic lecture. One factor possibly affecting reliability is the variation in quality of sound equipment used between test centres.

The 1989 version of the IELTS Speaking subtest assesses a range of language proficiency levels by adopting a guided, five-phase oral interview format. It was recognized that IELTS examiners worldwide would inevitably vary in their skill and level of training as interviewers, and therefore a carefully structured interview would increase reliability (Ingram, 1991: 191). The time allocated to the Speaking test, 11 to 15 minutes, was constrained by the need to reduce overall testing time compared with the ELTS, and for purposes of administration in test centres with large numbers of candidates. The five stages of the Speaking test, introduction, extended discourse, elicitation, speculation and attitudes, and conclusion, assess oral proficiency within the context of general knowledge and authentic situations relevant to the candidate. Although the interview format allows scope for some initiative by the candidate, particularly in phase three, the interviewer essentially retains control.

The Application of Bachman's Framework

Bachman's theoretical framework of factors influencing performance on language tests is discussed in Chapter Three. Within the framework, the categorization of test method facets is applicable to the analysis of a second language proficiency test in four ways: for describing the content of language tests and making comparisons between tests (Bachman *et al.*, 1988); for designing and developing new tests; for examining test validity; and for contributing to

research into the effects of test methods on performance (Bachman, 1990). However, in this study the framework will only be applied in describing the content of the IELTS test.

Bachman (1990: 119) classifies five test method facets: the testing environment, the test rubric, input, expected response, and the relationship between input and response. Since the IELTS is administered worldwide and in diverse situations, the testing environment, that is, the place of testing and the personnel, may be familiar for some test takers and unfamiliar for others. Furthermore, the physical conditions under which students are tested will vary unavoidably.

The facet of test rubric in the IELTS test has been discussed in detail earlier in this chapter. To summarize, the test is organized into four parts or subtests that are given different time allocations, and, since the Reading and Writing subtests are integrated to some extent, the sequencing of subtests is important. In addition, test takers have the opportunity to become familiar with test instructions by obtaining a specimen test paper. However, a recent administrator's report suggests that cultural factors may affect the ability of test takers to follow test instructions (IELTS Australia, 1994). For example, the instruction 'Write an essay for a university teacher' is open to cultural misinterpretation and may elicit the response 'Dear respected teacher'. The degree of 'speededness' as perceived by the test takers and the possible influence of testing time on performance are considered further in Chapter Five.

According to Bachman's definition of 'channel', the IELTS test includes both aural and visual input, and both oral and visual response. The sociolinguistic characteristics of the IELTS include the use of both Australian and British English dialects, variety of register, that is, formal written for the reading comprehension test and less formal for the listening comprehension test and oral interview, and authenticity of language. The fifth test method facet of Bachman's framework, the relationship between input and response, is generally

nonreciprocal in the IELTS, with the exception of the oral interview which is both reciprocal and adaptive.

Reporting Results

Brindley (1986) defines a proficiency rating scale as descriptors of stages of language behaviour along a continuum . A proficiency rating scale was used to report test scores on the ELTS test and nine bands were chosen for reasons of discriminating over a wide range of proficiency levels using general overall descriptions (Carroll, 1983: 98). The British Council favoured a continuation of this accepted and recognized system in a revised international test. Consequently, the IELTS test results are reported on nine bands of language proficiency which describe communicative and linguistic criteria from Non-User, Band 1, to Expert User, Band 9 (see Appendix II). A candidate's test report form shows results in two ways: first, as an average band score that gives an overall impression of competence; and, second, as a diagnostic profile of band scores on four macroskills, which reveals a test taker's strengths and weaknesses. For example, an overall IELTS score of Band 6 may be composed of 6.0 (Reading), 5.0 (Writing), 7.5 (Listening), and 5.0 (Speaking). The primary intention has been to enable test users to interpret IELTS test results easily.

Alderson (1991b) discussed in some detail the process of defining a nine-point proficiency rating scale for the new IELTS test. He outlined the procedure for converting raw scores on the indirect subtests of Reading and Listening into band scores. An explanation was also given of profile and global band descriptors for rating test takers on the direct subtests of Writing and Speaking. Alderson explained the process of directly converting a test taker's performance on each of these subtests to band scales from 1 to 9. Whereas it was decided to focus the ESP subtests of Reading and Writing around proficiency level Band 6, the subtests

of Listening and Speaking and the General Training module were designed to focus around Band 4 (Ingram, 1991: 193).

Since the General Training module is designed for candidates entering technical and vocational training courses, or senior secondary school, UCLES has set an upper limit of Band 6 on the Reading and Writing subtests for candidates taking this module. 'The logic, practical value and validity of this decision have yet to be fully tested' (Ingram, 1991: 189).

However, Brindley (1986: 22) pointed out that 'a general proficiency description cannot be regarded, nor should it be accepted, as a definitive *scientific* measurement of any given individual's language behaviour'. Proficiency level descriptors are not discreet and a test taker's performance may cross proficiency level boundaries. In addition, he suggested that proficiency rating scales tend to be 'impressionistic, subjective, relative in interpretation and therefore unreliable' (Brindley, 1986: 19). A number of research projects have been commissioned by IELTS to revise and validate bandscales for the Reading, Writing and Listening subtests, and to evaluate inter-rater reliability on the Speaking subtest (IELTS Australia, personal correspondence, May 1994).

An International Context

The IELTS test, although available in British Council test centres worldwide, was essentially designed for international students intending to study at British universities and colleges. However, the IELTS test is the result of a joint project between Australia and Britain, with some initial contribution from Canada, and consequently, the target situations widened. For example, the General Training module is particularly appropriate for the needs of the Australian ELICOS situation. In addition, the content of test items is no longer confined to British culture. The questions of cultural appropriacy and cultural bias were considered

carefully in the design of a new IELTS test (Ingram, 1991: 195). Whether the IELTS has been successful in its selection of issues and contexts shared by English-speaking cultures is an area of potential research. Furthermore, within the international arena the IELTS is a relatively new test and has some way to go before it is as widely established as the TOEFL.

Assessing an Innovation

According to Alderson (1986), the process of innovation in testing necessitates some fundamental questions. Why is there a need for a new test? From where does the need for change emanate? What evidence is there that the former test is ineffective? Do the test takers and test users perceive a need for innovation? It has been demonstrated in this chapter that by 1986 there was a generally held view amongst language testers and test users that the ELTS test was in need of revision. The Munby model, which had provided the construct on which ELTS was based, had been criticized as inadequate and outdated by applied linguists. The key issue in innovation was to search for an appropriate model of language proficiency for the development of a revised test. The Edinburgh Validation Study (Criper and Davies, 1988) inferred that the ELTS was effective as an ESP test; nevertheless, the model on which it was based was ineffective. Clearly the ELTS was established and accepted by test users, and any innovation would at least need to retain the level of face validity of the existing test.

The IELTS test was designed without a clear theoretical framework in a period of transition between the demise of the Munby model and the emergence of Bachman's framework for the design of language proficiency tests. As a consequence, although it is claimed that the IELTS 'reflects new developments in language testing' (IELTS, 1992: 2), it retains much of the legacy of the ELTS test which it replaced, and the test constraints imposed by the British Council. In terms of innovation in testing, it reflects evolution rather than revolution.

The IELTS Test 1995

Five years after its inception the IELTS test has undergone a further evolution. An updated version has been available for examination from April 1995. The revised test includes six major changes to the 1989 version which affect test takers. The description of these changes outlined below is based on the *Information Circular*, IELTS 1995 (UCLES, 1994a).

Firstly, although the macroskills-based structure of the test remains, the new IELTS test offers candidates a choice of two EAP modules rather than four for the Reading and Writing subtests, as shown in Figure 2.3. The ESP modules, A, B and C of the 1989 version, are replaced by one Academic module; however, the General Training module is retained.

Figure 2.3

The International English Language Testing System, 1995

(Deakin, 1994: 2)

<u>Listening</u> Comprehension questions (30 minutes)	<u>Academic Reading</u> Comprehension questions (60 minutes)	<u>General Training Reading</u> Comprehension questions (60 minutes)
<u>Speaking</u> Structured oral interview (11-15 minutes)	<u>Academic Writing</u> Task One (20 minutes) Task Two (40 minutes)	<u>General Training Writing</u> Task One (20 minutes) Task Two (40 minutes)

Secondly, test time is increased from 145 to 165 minutes. This represents an addition of five minutes for the Reading subtest and fifteen minutes for the Writing subtest. Despite an increase in time available to test takers, the required word count for the two writing tasks is also increased, from 250 words for the ESP modules and 200 words for the General Training module, to 400 words in total for each module, Academic and GT.

Thirdly, the Reading subtest for both the Academic and the General Training modules is divided into three reading comprehension sections, whereas this was not standardized between modules and versions in the 1989 IELTS test. There is no longer a link between the Reading and Writing modules; reading and writing were previously integrated.

Fourthly, there is a change in the order in which candidates take the subtests: Listening, Reading, Writing, Speaking. In the 1989 version of the IELTS test, Listening followed the Writing subtest. In addition, the Speaking subtest may be completed by candidates on a separate examination day, up to two days after the other subtests, 'at the discretion of the test centre' (UCLES, 1994a: 2).

In the 1989 version of the IELTS test, the General Training module was assessed over six proficiency rating scales or bands, and was considered to be less academically demanding. The 1995 version of the General Training module is comparable to the Academic module for both the Reading and Writing subtests. Both modules require two writing tasks of at least 150 words and 250 words, whereas 80 words and 120 words were previously required for the General Training module. Consequently, a fifth change is that Reading and Writing subtests for both Academic and GT modules are to be assessed over 9 Bands in the revised test. This increases the chances of test takers who choose the GT module scoring above Band 6 overall.

A sixth change which affects the test takers is an increase in the number of IELTS test versions to be made available to test centres in proportion to their number of candidates. Furthermore, test material is planned to remain in use for no more than one year. This should improve test security by reducing the likelihood of candidates facing recycled test material.

Proposed changes to the administration of the IELTS test will affect candidates indirectly and have implications for further research (UCLES, 1994a). Increasing numbers of international students are taking the IELTS test, particularly in Australia. The necessity for computerised administration at test centres should ensure continued rapid reporting of test results, test security and effective operation of the three month candidate waiting period between IELTS test retakes. The reliability of the subjective marking of the Writing and Speaking subtests is to be improved by double checking of borderline candidates. In future, all oral interviews are to be recorded to maintain test reliability.

A change in the process of data collection will enable more information to be gathered about test takers, item difficulty and discrimination, and the responses of particular cultural and linguistic groups to test items. The data bank should encourage further research into these areas and contribute to the preparation of an annual validation report (UCLES, 1994a: 5).

The possible implications for test takers of revisions to the IELTS test are considered in Chapter Six.

Concluding Note

The 1989 IELTS test was trialled on international students to obtain feedback on a revised test. It is suggested that it is important to consider test takers' perceptions of an existing test before the introduction of an innovation. In the light of changes to the IELTS test, it seems relevant and opportune to ask the test takers whether they perceive the current test to be

measuring what it purports to measure, that is, English language proficiency and readiness for academic study. An investigation into international students' perceptions of the 1989 version of the IELTS test is discussed in Chapters Four and Five. The issue of face validity, as well as other approaches to validating language tests, is first discussed in Chapter Three which is a review of the research literature in second language proficiency testing.

Chapter 3

Literature Review

Confusion, disagreement and uncertainty are reflected in much of the writing associated with language testing, a confusion which can be partly explained by a lack of agreement about the nature of language, language learning and use (Nunan, 1991: 47).

Introduction

This chapter is a review of recent research literature in second language proficiency testing. There have been major paradigm shifts in language testing that have reflected changes in the interpretation of language proficiency and in the terminology associated with it. The relevance of developments in English for Specific Purposes and communicative language testing to the design of language proficiency tests is demonstrated in this chapter. In addition, it includes consideration of a recently developed theoretical framework of communicative language ability. A summary follows of research into the effects of testing method on test reliability and validity and studies of test taker characteristics that affect test performance. Approaches to validating proficiency tests, particularly the IELTS test, are then examined. Finally, the importance of distinguishing between norm-referenced and criterion-referenced language testing is discussed.

Definitions of Language Proficiency

Nunan (1991) highlighted the confusion in terminology associated with language proficiency testing. He proposed that 'competence' refers to what a person knows about a language and

'performance' refers to what the person does with the language (Nunan, 1991: 47). He also stated that language proficiency is commonly held to be 'the ability to perform certain communicative tasks with a certain degree of skill' but argued that this definition is unsatisfactory (Nunan, 1988: 127). Nevertheless, Ingram (1985: 220) defined proficiency as 'not just knowledge but the ability to mobilize that knowledge in carrying out particular communication tasks in particular contexts or situations'. He also stated that defining language proficiency involves more than simply the ability to carry out tasks. A 'behavioural' definition of proficiency is needed to incorporate 'how' communication tasks are carried out (Ingram, 1985: 221).

Bachman (1989: 252) used the term 'communicative language ability' to include both competence (knowledge) and performance (a demonstration of the ability to use competence). Weir (1990: 7) supported this wider definition of language proficiency, but noted that it is difficult to distinguish between performance and competence in practice. Bachman's definition was based on the distinction between competence and performance originally proposed by Canale and Swain (1980). Bachman (1989: 251) stated that recent research studies in language testing tended to support the hypothesis that 'language proficiency consists of several distinct abilities that are related to each other or which are related to a higher order, general ability'. This hypothesis has replaced the 'unitary competence hypothesis' discussed below.

The psycholinguistic-sociolinguistic era in language testing in the 1970s was characterised by a trend away from discrete point tests towards global integrative tests, such as cloze and dictation (Weir, 1990: 3). Oller (1979, cited in Skehan, 1984: 202) argued that the structure of language proficiency consisted of one unitary factor, general or overall language ability, rather than a skills-based division of specific abilities. Furthermore, he asserted that overall language proficiency was most appropriately measured by integrative tests. According to Oller (1979, cited in Weir, 1990: 3), integrative tests, compared with discrete point tests,

more closely reflected the process of language use. However, Oller's Unitary Competence Hypothesis was subject to considerable criticism on both theoretical and empirical grounds applying principal components and confirmatory factor analyses (Hughes and Porter, 1983). Oller (1983) responded by conceding that he had overstated the theory of language proficiency as a single integrated trait, based on a 'pragmatic expectancy grammar' underlying language performance.

Controversy amongst applied linguists revived in the late 1980s. Does language proficiency consist of one unitary factor or is it divisible into separate and discrete factors, or alternatively, does it have both unitary and divisible aspects? However, research studies offered no clear-cut conclusions in favour of either the unitary or the divisible hypothesis (Buck, 1992: 315). The notion of a single psychological construct underlying language proficiency was replaced by the notion that language proficiency is 'multicomponential, consisting of a number of interrelated specific abilities as well as a general ability or set of general strategies or procedures' (Bachman, 1991: 673).

The issue of whether language proficiency consists of separate subskills or specific traits rather than a single unitary trait has implications for the design of second language proficiency tests. Multitrait/multimethod research design and confirmatory factor analysis were used in a number of studies to establish the construct validity of measures of communicative language ability. Palmer and Bachman (1981) found empirical support for the hypothesis that the subskills of reading and speaking are independently measurable. They also found evidence of a general language factor associated with communicative competence. Moreover, Buck (1992) demonstrated the existence of listening comprehension as a subskill independent of reading comprehension. Consequently, there appeared to be some empirical justification for skills-based language proficiency tests.

In addition, research suggested that the construct of language proficiency varies according to the level of proficiency (Alderson, 1991a: 12). At the lower and intermediate levels of second language acquisition, instructed language learners demonstrated differentiated proficiency on tests of macroskills (Milanovic, 1991: 135). Advanced learners, however, demonstrated more integrated language abilities or general proficiency, performing similarly in all macroskills. The test performance of beginners also tended to be unifactorial, rather than multifactorial.

The Purposes of Language Proficiency Testing

Firstly, the purpose of second language proficiency testing is to make judgements about the language abilities of test takers whose academic, cultural and linguistic backgrounds may be heterogeneous. Proficiency tests, unlike achievement tests, 'look forward rather than backward' and therefore predict future performance in diverse target situations (Alderson, 1988b: 17). Language proficiency tests are most frequently used to predict the readiness of non-native speakers of English to undertake English medium academic study. Together with an individual's previous academic records and profile, proficiency tests are also used to predict the probability of an international student being academically successful in the future target situation. Internationally recognized language proficiency tests, such as the IELTS and the TOEFL, claim to fulfill this purpose.

Secondly, language proficiency tests may be used for diagnostic purposes. A diagnostic profile of a test taker's language ability showing strengths and weaknesses can determine an individual's need for and the extent of further English language instruction. IELTS test results are presented as a diagnostic profile of the test taker's language proficiency on four macroskills (reading, writing, listening and speaking). In addition, results are also shown as an overall mean band score on the IELTS proficiency rating scale (see Appendix II). In order

to fulfill the two purposes of language proficiency testing, tests should be designed to measure language abilities that correspond to the communicative language abilities required in nontest language use, and test tasks should effectively simulate the target situation (Bachman, 1991: 681).

English for Specific Purposes

In meeting the goals of second language proficiency testing on an international scale, the question arises of how specific should proficiency testing be (Alderson, 1988b). The target situations of test takers will be diverse. In some respects, a practical solution is to produce one internationally recognized and standardized language proficiency test, such as the TOEFL test, for all test takers. Whilst the design and validation of one general proficiency test may be practical in time and money, is one test for all test takers fair and valid? (Alderson, 1988b: 18). Furthermore, do test takers perceive a general English language proficiency test to be acceptable, or do they perceive specific tests for specific academic purposes to have greater test appeal? Alderson (1988b: 18) commented that 'face validity, perhaps, argues for specific tests'; however, as face validity is unquantifiable, it is frequently given a low priority in the validation of second language tests.

The work of John Munby (1978) considerably influenced the development of English for Specific Purposes teaching and testing in the 1980s. His model of communicative syllabus design consisted of the Communicative Needs Processor and a detailed taxonomy of enabling skills. The former was a procedure for conducting needs analyses of second language learners and their target situations of communicative performance. Enabling skills are the underlying skills that enable students to operate in the target situation (Weir, 1990: 37). As Skehan (1984: 209) pointed out, the Munby model was essentially a taxonomy of 238 language subskills rather than an explanation of communicative competence underlying ESP

testing. Alderson (1988b: 19) argued that the content validity of a proficiency test depends not only on a needs analysis of diverse target situations, but also on analyses of the difficulties international students encounter in their academic studies and their perceptions of test appropriacy. Skehan criticized the Munby model for not providing an adequate framework to address the main difficulties which international students come across in academic study. He cited research by Johns and Dudley-Evans (1980, cited in Skehan, 1984: 210) which demonstrated that the comprehension of lecturers' styles and the content area of vocabulary knowledge were major difficulties faced by international students. Consequently, Skehan claimed that the Munby model had weaknesses when applied to ESP testing because it was primarily linguistic and did not 'take account of actual psycholinguistic processes that are important in language functioning' (Skehan, 1984: 211). For instance, the model did not take into account the effect of relevant cultural and background knowledge on text comprehension. In the period since the Munby model was devised, needs analyses of second language learners have become more complex and sophisticated (Johns and Dudley-Evans, 1991: 299).

Designing 'general' second language proficiency tests or English for Specific Purposes tests raises the issue of the relationship between background knowledge and language proficiency. Research has shown that previous background knowledge affects the performance of test takers on tests of listening comprehension and reading comprehension (Alderson, 1988b: 20). Therefore, it seems logical if the intention is to test general language proficiency that tests should be based on 'neutral' texts and tasks to avoid potential bias in favour of test takers with specific background knowledge described by Bachman (1990: 278):

Specific content will be interpreted as a potential source of test bias to the extent that individuals with background knowledge of the culture or subject content *systematically* perform better on the test than do individuals without such prior knowledge.

However, it is questionable whether completely neutral and unbiased texts are an achievable goal and whether it is valid to predict from general proficiency tests with 'neutral' texts to performance in specific target situations. On the other hand, it may not be valid to interpret from content specific tests to generalized knowledge (Bachman, 1990: 274).

'It is argued that appropriately differentiated tests in different skills areas need to be made available for evaluating different groups of examinees with different target situation needs.' (Weir, 1990: 11). However, in practice an infinite number of communicative language tests tailored to an infinite number of test purposes seems unworkable. Furthermore, distinguishing between generalized and specialized knowledge, and between general and specific language abilities is difficult (Bachman, 1990: 274).

Alderson (1988b: 18) raised the issue of parallelism or equivalent levels of difficulty between different academic modules in ESP testing. 'Ensuring parallelism through empirical, rather than descriptive, means is somewhat problematic.' (Alderson, 1988b: 28). This point is relevant to the 1989 version of the IELTS test, analysed in Chapter Two, which included parallel academic modules in three broad fields of study.

Communicative Language Testing

Communicative language testing is said to generalize to the target situation by testing within a meaningful communicative situation (Canale and Swain, 1980: 34). Weir (1990: 7) pointed out that:

In testing communicative language ability we are evaluating samples of performance, in certain specific contexts of use, created under particular test constraints, for what they can tell us about a candidate's communicative capacity or language ability.

Skehan (1988: 215) and Weir (1993: 167) both cited the work of Canale and Morrow to characterise the features of communicative language testing as follows: language is interaction-based and unpredictable; stimulus materials are relevant and authentic; language has a context; language is used for a purpose; discourse processing activities are realistic; and language is outcome evaluated. Weir (1993: 167) stated that 'we cannot build into our tests *all* the features of real life *all* the time. But we should maximise the involvement of these features'. Bachman (1991: 678) chose to focus on four characteristics to distinguish communicative language tests: the creation of an information gap requiring the processing of information from several sources of input; the interdependency of test tasks; the integration of test tasks and content within a given context; and the measurement of a broad range of language abilities.

Davies (1988b: 6) expressed the opinion that communicative language tests have no clear-cut definition. On the other hand, Weir (1990: 5) argued categorically that 'only direct tests which simulate relevant authentic communication tasks can claim to mirror actual communicative interaction'. However, Skehan (1984: 208) asserted that the inclusion of authentic communication tasks in a test can neither guarantee an adequate sample of communicative language use nor effective predictions of performance in the target situation.

As Shohamy and Reves (1985: 55) pointed out, the goal of interaction in a test is to assess the test taker's language performance and, hence, communication between participants is usually artificial. The examiner and the test taker are unlikely to interact in the same way in an oral proficiency test as in real life. For example, the topic of communication is imposed by the examiner. In addition, the physical environment is inauthentic and there are time constraints. Although some tests appear to test authentic language behaviour, particularly oral interviews, they are in fact simply assessing authentic 'test language' (Shohamy and Reves, 1985: 57; Stevenson, 1985a: 47). Spolsky (1985a: 39) explained:

Any language test is by its very nature inauthentic, abnormal language behaviour, for the test taker is being asked not to answer a question giving information but to display knowledge or skill.

Is the goal of authentic language testing unattainable, therefore? The authenticity of language testing is an area that has generated considerable research in the past decade and 'authenticity has been defined in a number of different ways' (Bachman, 1991: 689). First, there is the 'direct' approach. However, according to Bachman, this approach in language testing is unattainable since inferences about language ability must be based on indirect observations. Second, there is the real-life approach to authenticity. This is problematic when the target situations of test takers are diverse since language use in real life is varied and complex (Bachman, 1990: 312). Third, there is the face validity approach to defining authenticity. However, Bachman (1991: 690) claimed 'this definition refers to a purely subjective response on the part of the evaluator'. He proposed two different approaches: situational authenticity and interactional authenticity. Situational authenticity is 'the perceived relevance of the test method characteristics to the features of a specific target language use situation'; interactional authenticity 'is a function of the extent and type of involvement of test takers' language ability in accomplishing a test task' (Bachman, 1991: 690-691). Introspection techniques, discussed below, provide an understanding of test takers' strategies in completing test tasks and may help in assessing interactional authenticity (Cohen, 1984).

According to Weir (1988: 76), authenticity is probably an 'unattainable goal in mass language testing' but 'a *realistic* context is not'. For example, the Test in English for Educational Purposes (TEEP) was designed to integrate listening and reading texts with writing tasks in realistic contexts that simulate the target situation of academic study (Weir, 1990: 84). Weir argued, therefore, that integrated tests in EAP probably assess the construct underlying language proficiency more effectively than separate subtests of each macroskill. As Shohamy and Reves (1985: 52) remarked, authentic language testing implies the inclusion of both

linguistic and non-linguistic variables, and as a result, language performance can be complex, varied and difficult to control.

In Bachman's terms (1990: 317) test authenticity is essentially synonymous with communicative language use. He addressed the characteristics of the interactional/ability approach to authenticity by proposing theoretical frameworks of communicative language ability and test method facets. 'Two factors, the test taker and the characteristics of the test method, will jointly codetermine the extent to which a given test or test task is authentic' (1990: 318). In addition, the characteristics of the testing context were considered in the model.

A Model of Second Language Proficiency

Canale and Swain (1980) proposed a model of communicative competence consisting of three dimensions of knowledge and skills (grammatical, sociolinguistic and strategic competencies) rather than one global dimension as postulated by Oller. The model was later revised and expanded to include discourse competence. Within the framework, grammatical competence was defined as knowledge of the language code, sociolinguistic competence as the 'extent to which utterances are produced and understood appropriately in different sociolinguistic contexts', strategic competence as knowledge of verbal and non-verbal communication strategies, and discourse competence as coherence and cohesion (Swain, 1985: 37). The model was based on an assumption that learners develop competence in each area of knowledge and skills independently.

The application of psychometric theory to language testing was still dominant in the 1980s. However, criticism was being voiced of existing language proficiency tests based on norm-referenced testing. Bachman and Clark (1987: 21) commented that 'language testing

researchers are not working with a single paradigm, with respect to either a theory of language proficiency or a measurement theory'. They also questioned whether existing tests were measuring the range of language abilities necessary for communicative competence.

Skehan (1988: 211) reflected on the need for a theory of communicative language use:

Theory which guides and predicts how an underlying communicative competence is manifested in actual performance; how situations are related to one another; how competence can be assessed by examples of performance on actual tests; what components communicative competence actually has; and how these interrelate.

Bachman's interactional model of language test performance has probably been the most influential theoretical framework for language proficiency testing in the 1990s. According to Bachman (1991: 698), the framework enables language testers to 'specify and assess the relationship between language test performance and nontest language use'. The framework consisted of two components, language ability and test method facets, to 'make inferences about abilities or to make predictions about future language use on the basis of language test scores' (Bachman, 1991: 687). The first component, language ability, consisted of language knowledge (organizational knowledge and pragmatic knowledge) and metacognitive strategies (assessment, goal-setting and planning). The model originally incorporated the terminology 'language competence' and 'performance'. 'Competence' was later revised to 'knowledge' (Bachman, 1991: 683). The second component, test method facets or characteristics, has been discussed in Chapter Two with reference to the IELTS test.

Factors affecting language test performance within Bachman's framework are communicative language ability, test method facets, personal attributes, and random factors (1990: 165). Whilst Bachman acknowledged that random factors and personal attributes, such as sex, first language, background knowledge and field independence are variables which are generally beyond the control of language testers, he argued that communicative language ability and

test method facets can be controlled in test design and use. In fact, research has shown that test method effect, discussed below, is a significant factor in language testing (Alderson, 1991a: 11).

Test Method Effect

Test method refers to the procedure or technique that is used to assess the knowledge being measured in the test (Shohamy, 1984: 147). Research into test method effect on test performance has shown that test takers' familiarity with specific test methods may be reflected in differential language test scores. As Bachman (1990: 47) observed, language proficiency tests should use testing techniques which are familiar to students. Standardized tests using familiar methods, such as the multiple-choice test technique, are usually perceived by test takers and test users to have face validity, as defined in Chapter One of this study. However, test methods frequently become institutionalised. Consequently, there has been a reluctance to accept alternative test methods in second language proficiency testing because of the high reliability of established test methods (Alderson, 1991a: 8).

Discrete point test methods have been subject to criticism despite their objectivity and reliability. According to Hughes (1989: 60), the limitations of multiple-choice techniques are that they test only recognition knowledge, they encourage guessing, they restrict what can be tested and they tend to result in harmful washback effects. Traynor (1985: 45) argued that multiple choice items used in the TOEFL test encourage guessing. Skehan (1989) and Bachman (1990) cited a number of research studies that demonstrated how test methods affect language test performance. Research into the effects of test method can contribute to the design of fairer language proficiency tests that provide an opportunity for test takers to give their best performance by reducing test method effects (Bachman, 1990: 156).

According to Weir (1990: 11) inauthentic tasks result in a test method effect. However, as Palmer and Bachman (1981) demonstrated oral interviews, so-called 'authentic' direct tests, can also be subject to test method effects. Shohamy (1984: 149) stated that a major problem in language testing is to demonstrate method effect when measuring the traits of reading and listening comprehension which can only be observed indirectly.

Test Performance

Research has shown that language test performance can be affected by factors other than the ability being tested. Therefore the 'reliability of scores and the validity of their interpretations' may decrease (Bachman, 1990: 226). Skehan (1989) and Bachman (1990) cited research demonstrating that systematic differences in test performance were related to the individual characteristics of test takers, including linguistic and cultural subgroup, sex, age, field independence and rates of test anxiety. Lynch, Davidson and Henning (1988) used cluster analysis to study the implications of person dimensionality for language test validation. They commented that:

Elements such as native language background, age, sex, academic major and previous years of study in the target language may well provide important information in test validation in the sense that these elements may define significant examinee dimensions that affect test performance and the interpretation of test score variability (Lynch *et al.*, 1988: 206).

Chen and Henning (1985) identified vocabulary test items in an English language proficiency placement test which showed linguistic bias in favour of Spanish native speakers compared with Chinese speakers. Furthermore, the TOEFL test was strongly criticized for its cultural bias although intended to test the English language proficiency of international students planning to study in the United States. Traynor (1985: 45) argued that the TOEFL test is not

the place to test knowledge of American culture. If cultural or subject content is included in a language proficiency test and it is neither part of the ability intended to be measured by the test nor related to the purpose for testing, the construct validity of the test is doubtful.

Investigations into the effect of field independence on language test performance have proved inconclusive (Bachman, 1990: 276). Chapelle (1988: 66) referred to field independence or dependence as a 'variable of cognitive style' that affects 'the process of perception, thinking, and problem solving'. Her research into field independence/dependence as a learner characteristic related to test performance showed that field independent non-native speakers of English perform better on multiple-choice language tests. However, the results of the study were tentative. Allan (1992: 101) identified prior knowledge, text analysis, and test-taking skills as factors affecting reading comprehension test performance. He claimed that test-wiseness is differentially spread across student populations and affected the construct validity of tests. Test-wiseness is the ability of test takers to use test-taking strategies to enhance their test scores due to the characteristics and format of tests (Allan, 1992: 101-102). In summary, therefore, it may be necessary to consider the person dimension variables of international students when interpreting second language proficiency test results. As Henning *et al.* (1985: 142) argued:

Examinee performance is confounded with many cognitive and affective test factors such as test wiseness, cognitive style, test-taking strategy, fatigue, motivation and anxiety. Thus, no test can strictly be said to measure one and only one trait.

Research into the effect of background knowledge on test performance showed that international students performed better on reading comprehension tests in which the content area was based on their own specialist discipline (Alderson and Urquhart, 1985a; 1985b). Academic background seemed to affect test performance, but the effect was inconsistent and test method also appeared to be a variable affecting Alderson and Urquhart's research

(1985b: 203). However, test takers are generally at an advantage when taking a reading comprehension test in a familiar academic field, and at a disadvantage when taking a test in an unfamiliar academic field. Alderson and Urquhart (1985a: 41) suggested that prior background knowledge affected test performance by tending to 'top up' performance on language proficiency tests at a certain linguistic level. Reves and Levine (1992: 202) stated that 'well organized background knowledge has a facilitative effect' on text processing by helping the reader to make predictions about the text which are knowledge-based'. In addition, there is evidence that text type in listening comprehension tests affects test performance (Shohamy and Inbar, 1991).

The issue of test bias thus arises as a possibility when differential test performance occurs across groups of test takers and it cannot be attributed to differences in the language ability being tested. Consequently, the validity of inferences from test scores will be affected. There are a number of approaches to the gathering of evidence for the validity of test interpretations.

Approaches to Validity

The search for test validity is an elusive one ... the value of a test overall can best be estimated by a broad look at its stability, content, operational effectiveness, acceptability to all the users concerned and its educational effects (Carroll, 1983: 100).

Test validity is the extent to which a test measures the knowledge or abilities that it claims to be measuring. If a language test measures subject-matter knowledge or cultural knowledge that is not part of the linguistic or communicative competence that it purports to measure, the validity of the test is questionable. Validity may be defined as inferences made on the basis of test scores (Bachman, 1990: 243). The concept of test validity is also relative to the purpose of testing and the testing situation, although a test may be reliable for a particular purpose

without being valid. However, reliability and validity are closely related. According to Henning (1987: 91), the validity of a test is probably more important than its reliability.

Content Validity

Content validity is the extent to which a test adequately measures the specified skills and behaviours it purports to measure. To establish the content validity of a language proficiency test, the parameters of language proficiency should be clearly defined in the test specifications. Test items should sample the behavioural domain in proportion to the relative importance of the skills and abilities thus defined. Since testing time is usually a constraint on the comprehensiveness of language proficiency tests, sampling of the target domain is necessarily selective. Content validity depends on judgements by experts of the knowledge and skills which are 'comprehensive and representative of the content domain purported to be measured by the test' (Henning, 1987: 190). A study by Criper and Davies (1988) demonstrated that the content validity of the ELTS test was very doubtful. Consequently, subject specialists, language teachers, applied linguists and language testers were asked to judge the appropriateness of the test specifications and test items of the new IELTS in order to establish its content validity. The 1989 version of the IELTS test was judged by experts to be a relevant representation of English language proficiency for academic and vocational studies (Alderson, 1988: 226).

Weir (1990: 21) distinguished between qualitative *a priori* and empirical *a posteriori* approaches to validity. *A priori* content validity is particularly important when the purpose of a language proficiency test is also to obtain diagnostic profiles of test takers (Weir, 1990: 25). IELTS test results are reported as an overall band score and a semi-diagnostic profile which indicates a test taker's weaknesses in the four macroskills. Hence, the extent to which

an international student requires further English language instruction before commencing academic studies can be assessed.

Construct Validity

Content validity and *a priori* qualitative construct validity are closely related (Weir, 1990: 24). Construct validity is the extent to which a test measures the underlying learning theories or constructs that explain observable behaviour. Weir (1990: 23) stated that 'a test should always be designed on a principled basis, however limited the underlying theory'. Construct validity links theory and practice in testing and, according to Skehan (1984: 208), it is the most important validation of a language proficiency test. Clarifying the underlying theory governing a test is an essential prerequisite for establishing construct validity. Constructs 'are creations of applied linguists which, it is claimed, have some theoretical justification' (Skehan, 1984: 209). During the ELTS Revision Project, applied linguists were asked for their subjective judgements of the theoretical construct of language proficiency underlying the proposed new IELTS test. However, for reasons implied in Chapter Two, the experts failed to agree on an underlying construct of language proficiency for the IELTS. Consequently, establishing the construct validity of the test was undertaken by language testers and researchers. According to Alderson and Clapham (1992: 165), a pragmatic approach was adopted and 'thus the resulting test can only be at best an indirect operationalization of theory'.

A posteriori statistical validation procedures are also used to establish construct validity. The Cambridge-TOEFL Comparability Study investigated the construct validity of two internationally recognized standardized tests of English language proficiency (Bachman *et al.*, 1988). Although both tests measure language abilities and are designed for international students, the TOEFL test is strongly based on psychometric measurement theory, and the

Cambridge proficiency tests 'represent language test development guided primarily by applied linguistic theory' (Bachman *et al.*, 1988: 142). The TOEFL consists of three subtests: Listening comprehension; Structure and written expression; and Reading comprehension and vocabulary. There is also an optional subtest, the Test of Written English (TWE), and a separate Test of Spoken English (TSE). Bachman's theoretical framework of communicative language ability and test method facets, described above, was used to investigate the construct validity of the two tests. Henning (1989) was critical of the study on the grounds that the researchers attempted to compare an EAP proficiency test, the TOEFL, with a battery of tests of non-academic English, the Cambridge Certificate of Proficiency in English and the First Certificate in English, which were intended for different target populations.

Item Response Theory or latent trait measurement was used to establish the construct validation of communicative language proficiency tests (Adams *et al.*, 1987; Henning, 1984; McNamara; 1990). In trial tests undertaken in Australia and South East Asia, the IELTS was administered to a sample population of 3930 candidates, both native and non-native speakers of English, from diverse backgrounds and first languages. Griffin (1990) reported that the General Training module, which is of particular relevance in the Australian educational setting, demonstrated sound construct validity. The report verified that 83.3% of the items on the Reading subtest of the General Training module fit the underlying latent trait. Trial data on the general characteristics of the IELTS modules showed a wide range of item difficulty and discrimination, and high fit validity on a Rasch model.

Hamilton *et al.* (1993) investigated the performance of Australian native speakers on Module C exemplar versions of Reading and Writing subtests of the IELTS. Their research showed that the performance of native speakers was not uniform and could be related to educational level and work experience. They were unable to make a clear distinction between native speaker and non-native speaker performance on the IELTS test. Hamilton *et al.* found that native speaker subjects enrolled in Colleges of Technical and Further Education (TAFE) in

Victoria obtained a mean band score of 5.0 on the Reading subtest, that is, below the recommended entry requirement for international students intending to pursue academic studies. 'The results suggest that the IELTS Reading and Writing subtests are performance tests in the strong sense of the term.' (Hamilton *et al.*, 1993: 19). As the IELTS is a communicative EAP test simulating the performance tasks required of students in academic situations, their research showed that factors other than second language proficiency were also being assessed in the exemplar tests. The conclusions from this study were significant as validation of the IELTS criterion-referenced rating scales is essential for construct validity. If the native speaker is not the Expert User, defined by Band 9, then who represents the point of reference for the highest level of performance?

Criterion-Related Validity

Criterion-related validity is 'the extent to which test performance is related to some other valued measure of performance' (Hatch and Farhady, 1982: 251). The criterion-related validity of a newly designed test is usually established by correlating the test results of a sample of the target population with their results on existing well-established valid tests measuring an external criterion. However, there are two inherent problems in demonstrating criterion-related validity. First, in the process of test innovation there is the risk of validating a new test against its own parallel forms, rather than validating it against an independent external criterion (Baker, 1989; Bachman, 1990: 249; Davies, 1983). Second, although criterion-related validity demonstrates that a test can predict performance on a specified criterion, correlational analysis assumes a normal distribution of test results. As Brown (forthcoming: 423) points out, criterion-referenced tests often yield a skewed distribution of results that cannot be reliably interpreted in terms of criterion-related validity.

Criterion-related validity may be classified into predictive validity and concurrent validity. In language testing, the two types of validity are usually distinguished on the basis of time. Concurrent validity is concerned with the correlation between test scores and an available criterion measure obtained at approximately the same time, whereas predictive validity is concerned with the correlation between test scores and a criterion measure obtained at a later date. Anastasi (1982: 137) clarified the distinction between the objectives of concurrent validity, that is, the diagnosis of existing status, and predictive validity, or the prediction of future outcomes.

During the test trials undertaken in Australia in 1988, data was gathered on correlations between the IELTS subtests and criterion measures, such as the existing ELTS test, the TOEFL test, the Hong Kong GCE Examination results, and the Oxford tests. The results showed generally high correlation coefficients between the IELTS and other tests of language proficiency (Griffin, 1990). The importance of establishing the criterion-related validity of language proficiency tests which are used to make crucial decisions about the acceptance of test takers into overseas tertiary institutions was emphasized by Weir (1990: 28).

Predictive Validity

Predictive validity is the extent to which a test accurately forecasts a test taker's future performance in the target situation. For example, 'a university admissions test is said to have predictive validity if its scores correlate highly with a performance criterion such as university grades' (Henning, 1987: 196). Considerable time and effort in research continue to be devoted to the predictive validity of language proficiency tests. Light *et al.* (1987) found a weak correlation between TOEFL scores and grade point average (GPA) in their study of international students at an American university, thus concluding that 'merely knowing how a student scored on the TOEFL test will tell us practically nothing we need to know to predict

the student's academic performance' (1987: 255). Their study showed that the relationship between TOEFL and GPA scores varied according to international students' subject specialisms. They also found that students with TOEFL scores below a 550 cut-off point were as academically successful as students with scores above 550 TOEFL. Graham (1987) reviewed a number of academic prediction studies to determine the relationship between academic success and English language proficiency. However, she found that research conclusions were contradictory, and the relationship between English language test scores and academic success was tenuous. In addition, studies showed that test scores and future academic success do not always correlate highly, implying that other factors contributed to success or failure.

There are many variables affecting the predictive validity of a language proficiency test before the performance criterion of academic success is assessed. These include the extent to which an international student interacts with native speakers, subject-specific background knowledge, positive or negative attitudes towards the host culture, motivation and other affective factors, and attitudes to learning (Burns, 1991; Graham, 1987). The factors which international students perceive to be of importance in academic success are discussed further in Chapter Five.

When considering predictive validity, it is important to be aware that research studies used a variety of criteria to measure academic success in the target situation and a variety of second language proficiency test results. Language proficiency is often defined by test performance on different types of tests; therefore, correlations between scores on multiple-choice tests and performance tests, for example, must be interpreted cautiously. Moreover, comparisons between studies are often complex because of the many variables that affect the results of these studies, including sample size, first language, English as a Second Language (ESL) or EFL context, subject specialization, and undergraduate or graduate level study (Graham, 1987: 508).

The IELTS test aims to provide a measure of the English language proficiency of international students and to demonstrate their readiness to undertake a course of academic or vocational study in the medium of English. Gibson and Rusek (1992) conducted research in South Australia into the predictive validity of the IELTS test by comparing test takers' overall IELTS band scores with their subsequent performance on tertiary level courses of study. Their research, based on a sample of 63 international students, demonstrated no significant relationship between overall band score results on the IELTS test and academic success in the first semester of university study. However, academic success was judged to be successful completion of the first six months of academic study, whether undergraduate or postgraduate, within a number of different university departments in. In other words, according to Gibson and Rusek's study (1992: 48):

If the student has fulfilled the requirements of the academic work and has been permitted to pass on to the next phase of the study program, she or he is deemed to have been successful.

Elder (1993) investigated the relationship between the language proficiency levels of international students as measured by the IELTS test and their subsequent performance in postgraduate teacher training courses. She concluded that in the short term the overall IELTS band score predicted academic performance in these courses to a limited extent. However, the criterion measure of academic success varied between the six different educational institutions in the study. In the longer term, many other variables again appeared to intervene to affect academic success, such as interpersonal skills, subject knowledge, cultural adaptability and motivation, thus diminishing the predictive relevance of language proficiency scores (Elder, 1993: 75). Elder (1993: 83) claimed that language proficiency thresholds and the demands of particular academic courses were difficult to balance, and consideration should also be given to the level of English language support that is available to international students during their studies.

Establishing the predictive validity of a standardized English language proficiency test, particularly if it is to gain international recognition, is in the best interests of the testing service or examination board. Consequently, cut-off scores on language proficiency tests used to select international students at different tertiary institutions and studies of predictive validity demand careful scrutiny. Prediction studies understandably do not assess the academic potential of international students with low levels of language competence who have failed to gain admission to universities, or who are undertaking further English language instruction to attain the threshold level. Academically successful students with generally high levels of language proficiency are admitted to overseas institutions and they represent the subjects of academic prediction studies.

Face Validity

Face validity is 'the degree to which a test appears to measure the knowledge or abilities it claims to measure, based on the subjective judgment of an observer' (Richards, Platt and Weber, 1985: 102). Anastasi (1982: 136) explained more explicitly who the observer might be:

Face validity pertains to whether the test 'looks valid' to the examinees who take it, the administrative personnel who decide on its use, and other technically untrained observers. Fundamentally, the question of face validity concerns rapport and public relations.

Face validity is thus based on the judgements of test takers and test users, not on the judgements of experts. When applied to the validation of a standardized English language proficiency test, face validity may be defined as the extent to which the test and its format seem to be relevant to the test takers in content, relevant to performance in the target situation, and culturally appropriate. For many reasons it is advantageous for a language proficiency test to have high face validity. According to Nevo (1985: 288), research showed that high

face validity may motivate test takers, attract candidates to sit for a test, reduce feelings of unfairness amongst test takers who perform poorly, and encourage the acceptability of a test.

During the trialling of the IELTS test, test takers were consulted about their perceptions of the new test and also asked to complete a self assessment rating of language ability on a nine point band scale (Griffin, 1990: 5). Data on test perceptions was gathered by asking six closed questions for each subtest:

- Do you feel that this was a fair test of your English?
- Was there enough time for you to complete the test?
- Was the test too hard?
- Was the test too easy?
- Were the questions realistic?
- Were the instructions clear?

These questions only offered a limited insight into the test takers' perceptions of the IELTS and Griffin's report on the Australian test trials dealt but briefly with their responses. The possibility of designing effective questionnaires to assess test takers' perceptions of language proficiency tests is discussed in Chapter Four.

According to Henning (1987: 96), response validity is the extent to which test takers respond in the way expected by the test designers. Responses may be invalidated by inadequate instructions or lack of familiarity with tasks, for example, or a negative attitude towards a test. In this respect face validity and response validity appear to overlap. Skehan (1984: 208) noted renewed interest in face validity or 'acceptability to the test taker' with a paradigm shift towards communicative language testing. However, Stevenson (1985a) warned against too much faith in the face validity of communicative language tests at the expense of empirical validity. He emphasized that there is no inherent validity in so-called authentic language tests; the testing situation is by definition inauthentic. This view was supported by Shohamy and

Reves (1985: 57) who asserted that the so-called real-life language of authentic tests is simply authentic 'test language'.

Skehan (1984) added a note of further caution. He observed that test takers' views of face validity are seldom constant. 'Face validity should therefore be concerned only with the extent to which the test format does not interfere with the deeper aims of a test' (1984: 208). Savignon (1983: 237) claimed that face validity is not an issue and although important to test takers, it has no empirical basis and is therefore ineffective for test validation.

However, as Weir (1990: 26) commented, if test takers are to perform in a way that reflects their language ability, they must find the test acceptable and fair. From this perspective, face validity should not be considered as merely 'cosmetic'. In Bachman's words (1990: 288-289):

The 'bottom line' in any language testing situation, in a very practical sense, is whether test takers will take the test seriously enough to try their best, and whether test users will accept the test and find it useful. For these reasons, test appearance is a very important consideration in test use.

Washback Validity

Morrow (1986: 6) introduced the term 'washback validity' to express the extent to which a test positively or negatively influences teaching and learning. This approach to test validity describes the relationship between testing and the teaching and learning that precedes it in the classroom (Weir, 1990: 27). However, Alderson and Wall (1993: 117) emphasized the need for more research into the nature of washback, and how it operates negatively or positively, since the concept is not clearly defined. Even a poor test may have a positive washback effect

on student motivation. Factors such as teacher competence and quality of educational resources may intervene and affect washback validity. Consequently, 'the quality of the washback might be independent of the quality of the test' (Alderson and Wall: 118).

The TOEFL test has been criticized for its negative washback effects since it primarily tests the receptive skills of reading and listening. Therefore, speaking and writing tend to be neglected in TOEFL preparation classes although there is an optional test of writing (TWE) and a separate, stand-alone test of speaking (TSE) administered by the Educational Testing Service (Traynor, 1985: 44). Another negative washback effect is that the test may encourage rote learning of vocabulary and grammatical structures. Furthermore, research has shown a significant test method effect, particularly in the case of multiple-choice tests (Shohamy, 1984). Test takers may become very adept at taking the TOEFL test, and since there is no penalty for choosing a wrong answer from the multiple-choice test items, the strategy of guessing is used by some international students. In contrast, the IELTS test encourages positive washback effects by testing performance on communicative tasks which simulate the target situation. Both teaching and learning of the four macroskills are necessary for the test, and test preparation is also preparation for the target situation. The washback effects of the IELTS and the views of international students about test preparation are issues raised again in Chapter Five. Communicative tests of English for Academic Purposes, which contextualize language proficiency and include a representative sample of tasks from the behavioural domain, are more likely to have positive washback effects since they prepare the test takers for the target situation of academic study (Hamilton *et al.*, 1993: 20; Weir, 1993: 97).

Questionnaires and interviews can be used to investigate the washback effects of external tests on classroom practice, but direct observations in the classroom provide a more reliable avenue of research. Hughes (1988) reported on the powerful beneficial washback effects on the curriculum of a new second language proficiency test introduced at an English-medium university in Turkey. A study in Sri Lanka by Wall and Alderson (1993, cited in Alderson

and Wall, 1993: 126) showed that innovation through testing had little effect on the way teachers teach and only affected what they teach. The introduction of a new version of the IELTS test in 1995 seems to offer an opportunity to monitor internationally the washback effects of the test. As Carroll (1983: 100) stated 'no tester in the educational field can afford to ignore the effects of his tests on the processes of education itself'.

A closer examination of the 'unduly simplistic' Washback Hypothesis is needed, as 'washback, if it exists—which has yet to be established—is likely to be a complex phenomenon which cannot be related directly to a test's validity' (Alderson and Wall: 116). A test will influence what teachers teach and what learners learn, how teachers teach and how learners learn, the rate and sequence of teaching and learning, the degree and depth of teaching and learning, and attitudes to the content and method of teaching and learning (Alderson and Wall: 120).

Introspective Methods

There has been an interest in research into the role of introspective methods in test validation (Faerch and Kasper, 1987). Cohen (1984) investigated the strategies and mental processes used by second language students to take reading tests. From his research findings he concluded that 'students may not be displaying a representative performance of their language competence' in their test responses, due to weaknesses in test format or in test-taking strategies (Cohen, 1984: 71). He proposed that either test takers should be trained in the techniques required by particular test methods or that the test methods should be changed.

The validity of indirect tests that encourage test takers to develop proficiency in particular testing methods, such as the cloze or multiple-choice, is nevertheless questionable. Shohamy (1984: 147) claimed that 'one characteristic of a good test is that the method has little effect

on the trait'. Test results which reflect ability in the testing method rather than the trait being measured have low validity. Her research showed that the test methods used in reading comprehension tests for second language students affected test results, especially at lower proficiency levels. Shohamy also proposed further research on introspective procedures to understand the strategies that test takers use and to develop testing methods that effectively measure the reading comprehension trait. As a result, the validity of language proficiency test results may improve. Weir (1990) suggested that thinking-aloud and retrospection techniques may be used to check 'expert' judgements of test validity. Alderson (1991a: 21) proposed that collecting data on test processes through introspective methods provides a valuable insight into a 'student-centred perspective on test validity'. The feasibility of using introspection methods to gain further insight into students' perceptions of the IELTS test is explored in Chapter Six.

Summary of Approaches to Validity

The role of various approaches to test validity is open to interpretation. In proposing that more attention be given to concurrent validity, Davies (1983: 145) commented that 'external validation based on data is always to be preferred to armchair speculation'. Nevertheless, he later concluded from validation studies of three English language proficiency tests that concurrent and predictive validity only served to establish the accuracy of content and construct validity (1984: 50), and that construct and content validity were the most important ways of validating tests of language proficiency.

Grotjahn (1987: 71) commented that 'to consider mainly statistical properties in test validation, is very problematic'. Validating a new proficiency test on criterion measures, usually other validated tests, also depends on their established construct validity, and thus a circulatory process is set in motion. Consequently, Grotjahn reiterated the need for qualitative

introspective methods as a means of validating the construct of language proficiency. Other researchers have also drawn attention to the problems of validating communicative tests against criterion measures of indirect, non-communicative tests (Weir, 1990: 28). Weir also stated that, within a sociolinguistic-communicative paradigm, more attention should be given to both content and face validity.

Heaton (1988: 160) argued that 'most designers of communicative tests regard face validity as the most important of all types of test validity'. However, Bachman (1990: 285) cited Mossier, Cronbach and Stevenson, amongst others, who rejected the significance of face validity. Bachman (1990) categorically stated that 'test appeal' is not a sound basis for validation. He also offered a timely reminder that 'while the relative emphasis of the different kinds of evidence may vary from one test use to another, it is only through the collection and interpretation of all relevant types of information that validity can be demonstrated' (1990: 237). In addition, he identified construct validity as the unifying concept for content and criterion-related validity (1990: 256).

The researcher's personal view is that research into the face and washback validity of language proficiency tests should not be overlooked. 'Expert judgements' of standardized tests are only one approach to improving test validity and as Alderson (1991a: 22) stated, 'research suggests that it may be unwise to leave *expert* judgements unchallenged'. Crucial selection decisions continue to be made about international students based on their language proficiency test results. Testing significantly influences educational and social decisions made about students (Bachman, 1990: 237). As long as language proficiency tests are used internationally for a variety of purposes, test takers and users will continue to judge the relative merits of each test. Consideration is given to face validity and washback effects in language proficiency testing in Chapter Five. The findings of a small-scale survey of international students' perceptions of the IELTS test are described. As Weir (1990: 27) remarked:

A suitable criterion for judging communicative tests in the future might well be the degree to which they satisfy students, teachers and future users of test results, as judged by some systematic attempt to gather quantifiable data on the perceived validity of the test.

Norm-Referenced and Criterion-Referenced Testing

The distinction between the two frames of reference for interpreting language test results, norm-referenced measurement and criterion-referenced measurement, dates back to the 1960s and the work of Glaser (Popham, 1978). Although norm-referenced testing (NRT), reflecting the 'psychometric-structuralist' period of language testing, tends to dominate the field of second language testing, some researchers have given prominence to the role of criterion-referenced testing (CRT). Bachman (1989: 248) claimed that criterion-referenced testing 'has been virtually ignored by language testing researchers and writers of texts on language testing'. This viewpoint was supported by Brown (1990: 78) who noted the dearth of criterion-referenced approaches in the field.

A norm-referenced test is 'designed to measure how the performance of a particular student or group of students compares with the performance of another student or group of students whose scores are given as the norm' (Richards *et al.* 1985: 68). Norm-referenced testing 'usually implies standardization through prior administration to a large sample of examinees' (Henning 1987: 194). Thus, the 'norm group' samples the test takers who the test is designed for, in other words, the target population. 'A norm-referenced test is designed to ascertain an examinee's status in relation to the performance of a group of other examinees who have completed that test'. (Popham, 1978: 24).

Brown (forthcoming) has emphasized that norm-referenced tests are most appropriate for measuring general language abilities or proficiencies. The TOEFL test is an example of a standardized, norm-referenced second language proficiency test that measures the general or global language abilities of international students. Norm-referenced testing is intended to disperse the test takers' scores along a normal distribution curve which represents a continuum of language proficiencies. A test taker's score is interpreted with reference to the mean score and standard deviation of the norm group. The percentile score shows the individual's relative ranking in comparison with other test takers. For example, based on the scores of examinees who took the TOEFL test between July 1991 and June 1993, a TOEFL score of 500 indicated a percentile rank of 38 (ETS, 1994: 30). Norm-referenced testing aims to discriminate between individuals in terms of proficiency by maximizing their score differences.

In contrast, a criterion-referenced test 'measures a student's performance according to a particular standard or criterion which has been agreed upon' (Richards *et al.*: 68). A criterion-referenced test is representative of a specified criterion level of ability or a well-defined content domain (Bachman, 1990: 74). It is 'used to ascertain an individual's status with respect to a well-defined behavioural domain' described in the test specifications (Popham, 1978: 93). Absolute performance, rather than relative performance, is measured and interpreted in terms of prespecified performance standards. According to Hudson (1991: 160):

Criterion-referenced language tests (CRTs) are generally designed to distinguish between two or more defined levels of language ability rather than to determine the status of individuals relative to other individuals who have taken a test.

Brown (forthcoming) has suggested that criterion-referenced testing is most appropriate for achievement and diagnostic tests. However, Bachman (1989; 1990) proposed a framework

for developing and using criterion-referenced tests of language proficiency in language programme evaluation.

Critics have pointed out that norm-referenced testing does not meet the needs of English for Specific Purposes testing as language proficiency is not measured in relation to a particular target situation. However, test users sometimes interpret scores on norm-referenced tests as if this were the intention. Judging by the minimum score required on the TOEFL test, usually 550, for entry to universities in Australia, an international student with a score in the 60th percentile on this standardized test is linguistically ready to undertake an academic course. In fact, critics of NRT have claimed that norm-referenced tests give little indication of the language proficiency of a test taker, except in relation to others who sat the test at the same time (Bachman, 1989). In addition, norm-referenced tests give little indication of the level of mastery of specific language content or skills. They 'fail to provide a lucid description of what an examinee's performance actually signifies' (Popham, 1978: 76).

There are a number of difficulties with CRT. One problematic concept is that of a cut-off score to define a minimum competency for certain decision-making purposes (Hudson, 1989). Test takers' scores on the IELTS test are reported as band scores on a descriptive proficiency rating scale of criterion levels of language proficiency (see Appendix II). There is effectively no pass or fail score, or cut-off point. However, tertiary institutions in Australia usually require a minimum level of proficiency of Band 6 on the IELTS test before international students are considered linguistically ready to undertake academic studies, as discussed in Chapter Two. A second problem is defining language proficiency with reference to the native speaker. Bachman (1989: 254) argued that native speaker proficiency is not an adequate criterion for language proficiency because native speakers vary in language proficiency, they vary in their dialects, and the concept of the native speaker is an abstraction. According to Hamilton *et al.* (1993: 7), the native speaker is not referred to explicitly in the IELTS proficiency rating scale, but is implied in the descriptor of Expert User, Band 9 (see

Appendix II). They claim that there has been a lack of research into the performance of native speakers on communicative language proficiency tests and 'reference to the native speaker as some kind of ideal or benchmark in scalar descriptions of performance on performance tests is not valid' (1993: 20).

Hughes (1986: 33) stated that the main benefit of criterion-referenced testing is the washback effect. Performance on criterion-referenced tests is directly related to the skills required in the target situation, and test preparation is therefore preparation for the target situation. Hughes also claimed that high face validity is another benefit of CRT if the rationale of the test is made clear to test takers and test users, and if the test effectively samples the language skills of the target situation. Hughes disagreed with Skehan (1984: 215) that 'criterion-referencing is only possible when the target behaviour is specified with great exactness ' and therefore too difficult to achieve in practice. He supported his argument with the example of a criterion-referenced EAP test designed and implemented in an EFL setting (Hughes, 1988b). In summary, therefore, criterion-referenced testing is subject to various interpretations and approaches (Bachman, 1988; Hudson and Lynch, 1984; Hughes, 1986).

Summary

The definition and interpretation of the nature of language proficiency depend on the purposes of testing. Therefore, language proficiency test innovations should develop within a framework that takes into account the target situation, the communicative context, and the theoretical construct of language proficiency (Weir, 1993: 169). Furthermore, 'the most important quality to consider in the development, interpretation, and use of language tests is validity' (Bachman, 1990: 289). Investigating face validity has been a neglected area of research in language proficiency testing. How can it contribute to the goal of improving 'the match between teaching, testing and target situation' (Weir, 1993:169)?

Chapter 4

Research Methodology

Future innovations in testing should perhaps pay more attention to the students' own informed view on assessment and on the methods which will enable them to perform to the best of their ability (Alderson, 1986: 99).

Introduction

As noted above, research in EFL testing has tended to neglect consulting the 'test-takers' about their perceptions of language proficiency tests, their attitudes towards the testing experience and their individual test performance (Alderson, 1986: 99). The aim of the research survey was to investigate international students' perceptions of the International English Language Testing System test. This chapter is a description of the research methodology used to carry out the investigation. The research findings will contribute to an analysis of the IELTS test, particularly with reference to the issues of face validity and washback validity discussed in Chapter Three.

Background to the Investigation

It was noted in Chapter One that before studying in Australia, many international students are required to sit a standardized English language proficiency test, the IELTS or the Test of English as a Foreign Language, as a measurement of their language ability. Both internationally recognized tests measure the second language proficiency of non-native speakers of English, yet they are very different in appearance and ideology. These tests affect

judgements with far-reaching consequences made about thousands of international students every year and their admission into Australian tertiary institutions. In 1994, the IELTS test was in the process of a major revision in appearance and, to some extent, ideology, effective from April 1995. It seemed an appropriate time to examine the perceptions held by test takers of the current 1989 version of the IELTS test; in other words, to find out the opinions of those international students who had taken the IELTS test previously and who were studying at an Australian university. In addition, it was also intended to make some comparisons between test methods, testing experience, test preparation and relevance to the target situation of the IELTS and the TOEFL tests based on the perceptions of international students who had taken both tests. It was hoped that students would find the subject of research both relevant and interesting, and therefore feel motivated to take part in the survey and follow-up interview.

Subjects of Research

The subjects chosen to take part in this survey were international students studying at the University of Tasmania, Launceston campus. Since the international student population studying at this campus is relatively small and accessible, it was initially decided to conduct a census survey of all international students. Approximately 270 students were on record in the International Student Office at the time of this study (March 1994). 72% of those students were from Indonesia, Malaysia, Singapore and Thailand; 9% from China and Hong Kong; 6% from India; 5% from Japan and Korea; and the remaining 8% were from other countries. Students were studying at undergraduate and postgraduate levels in the Faculties of Business, Design, Education, Engineering and Surveying, Humanities and Social Sciences, Medicine and Pharmacy, Nursing, Science and Technology, and Visual and Performing Arts. Some students were taking courses in the English Language Centre.

Instrumentation

Two approaches were chosen for data collection: a questionnaire and an interview. The main instrumentation for the sample population was the questionnaire, with a follow-up focus group interview.

Questionnaire

A questionnaire seemed the most appropriate method for sampling the total population within a short time and without major expense. Since the subjects of the research were primarily non-native speakers of English, it was judged that a questionnaire, rather than an interview, might result in a higher response rate. However, respondents were not given the opportunity to answer questions in their first language. A copy of the questionnaire is available in Appendix VI.

For several reasons, closed questions, in which respondents select from pre-set options, were chosen for the survey in preference to open questions. Firstly, respondents did not have to worry about their grammar skills when answering closed questions and the response rate amongst non-native speakers of English might be higher. Consequently, most written information on the questionnaire was limited to the type of biographical data frequently requested by a university's administration: nationality, first language, age, arrival date in Australia, and course of study. However, opportunity was provided on the questionnaire for additional comments. Secondly, closed questions might act as cues for longer-term memory, especially for students who took the IELTS test over one year ago. Open questions are probably more demanding in terms of recall. Thirdly, open questions may lead to frustration if the respondent is unsure about the type of answers expected whilst closed questions compel the respondent to choose only from alternatives selected by the researcher. A fourth point in favour of closed questions is that quantification of data is simplified as answers will vary more when open questions are used and the procedure is probably more time-

consuming for the respondent. Finally, the answers to open questions may be incomplete, they may digress from the topic and non-native speakers may have difficulty in expressing their opinions.

Interview

The objectives of the interview were twofold. Firstly, it was to serve as an information gathering process rather than a measurement survey. The goal was to obtain in-depth opinions and to probe interviewees' perceptions. The role of each interviewee in this type of research approach becomes that of an 'informed collaborator', as defined by Fowler and Mangione (1990: 19). It was particularly hoped that a comparison could be made between students' perceptions of the IELTS and the TOEFL tests. The structure of the interview was therefore more open than that of a standardized interview, with scope for interviewee initiative and for recording narrative type answers. Secondly, the interview was to serve as a follow-up to the questionnaire. The purpose was to attempt questionnaire validation by checking and clarifying the concepts used.

It was decided to conduct a group interview to be recorded on cassette and transcribed. Research has shown that focus group interviewing as a qualitative research technique can be effective as a method of investigation and aids the interpretation of results from surveys (Bers, 1987; Brodigan, 1992). A focus group interview usually involves small numbers of subjects from the target population in which the researcher is interested. The interactive nature of group interviews encourages interviewees to act as stimuli to each other and the researcher may be able to identify shared meanings. One group member's contribution might elicit another group member's contribution on the same topic (Carey, 1994: 229). The results of group interviews represent a source of social knowledge, in this case the perceptions of international students about two tests of English language proficiency, which may differ from the sum of the perceptions of individuals. In addition, it was felt that a group interview would help non-native speakers express their opinions more confidently. In terms of recall it

is preferable for interviewee to be prompted by interviewee, rather than by interviewer. A copy of the interview format is available in Appendix VII.

Pretesting the Questionnaire

Once the questionnaire was designed, a pilot study was undertaken to reveal possible deficiencies in instrumentation. The subjects chosen for pretesting the questionnaire were international students studying in the English Language Centre (ELC) on the Launceston campus. The researcher had previously taught in the ELC and had some contact with most of the students. Furthermore, the Director of the ELC and language teachers were supportive of the research project. The choice of subjects seemed appropriate as the researcher expected the English language skills of this group of international students to be less proficient than those of the wider international student population enrolled on academic courses. Consequently, piloting the survey was designed to identify lexical problems, confusing instructions, difficulties with specific questions and possible misunderstandings.

At the time of the pilot study, twenty-two students were attending English classes. These subjects represented a random sample of the subgroups of the target population. Some students were attending an intensive pre-sessional English class before entering undergraduate or postgraduate courses. Some students were studying English before taking an IELTS test and applying to Australian universities or colleges. Others were following a general English course at the university before returning overseas. The aim of the research was explained to the subjects. They were asked to participate voluntarily and the issue of confidentiality of information was discussed. All twenty-two students were willing to participate in the pilot study.

The pilot study included the questionnaire, an accompanying cover letter, a statement of informed consent, and an information sheet explaining the purpose of the study. Group discussion during a pilot study is one way to establish a clearer understanding of how subjects view the terms and concepts in a questionnaire. Therefore, the researcher was present when the survey was conducted and was able to obtain immediate comments from respondents that highlighted both major and minor problems in instrumentation. It was also possible to estimate the average amount of time needed by a respondent to complete the questionnaire.

Some terminology in the letter and information sheet accompanying the questionnaire proved to be a problem for respondents and needed clarification, such as *subject* of research, *anonymous* and *confidential*. Revisions were made accordingly for the final versions (see Appendices IV and V). The format of the information sheet was presented as a series of questions and answers to assist comprehension.

The questionnaire was redesigned as a result of the pilot study (see Appendix VI). Thirteen of the twenty-two respondents had taken an IELTS test. Initially it was intended to collect data from international students who had not taken an IELTS test, but intended to do so in the future. However, the pilot study showed that different instructions for past test takers and future test takers were confusing, and only a limited amount of useful data could be gathered. Also, future test takers were being asked for information about a test they had not taken and therefore could not judge. Several questions needed clarification as a result of the pilot study which served to highlight that 'the meaning of questions and response options must be made clear to respondents if their answers are to be properly interpreted' (Foddy, 1993: 161). The goal was to design specific, brief and grammatically simple questions.

Definitions of unfamiliar terms were included in the revised questionnaire as a result of pretesting. The necessity to provide additional space for comments was also apparent to

allow for responses not anticipated by the researcher. In addition, the pilot questionnaire included the descriptors of language ability used in the IELTS band score results, such as competent and modest (see Appendix II). However, during pretesting most respondents had difficulty with these terms. Unfortunately, the design of the survey made it impossible to avoid evaluative words altogether, such as *good*, *fair* and *adequate*.

To help recall, question 1 in Part 6 was revised to include the amount of testing time given to each macroskill in the IELTS test. Since research shows that many respondents give substantive answers when pressurized by the lack of a non-substantive filter (Foddy, 1993), the category 'not sure' was added to the final version of the questionnaire (see Appendix VI).

Pretesting the Interview

A pilot interview was conducted with two volunteer interviewees and recorded on cassette. Due to the limited sample of questionnaire respondents who had taken both the IELTS and TOEFL tests which were the criteria for interview, the task of piloting was problematic. The pretest sample was too small to identify all potential interview errors. However, the procedure helped in a small way to reveal ambiguous questions requiring clarification or elaboration. The pilot study also assisted in developing a strategy for conducting a group interview. Consequently, an interview format was carefully designed and copies were made available for each interviewee (see Appendix VII). This was an attempt to reduce test bias by ensuring that each interviewee in the focus group was presented with the same questions both orally and in writing.

Design

Questionnaire

Part 2 of the questionnaire was intended to record biographical data with the objective of providing a profile of the subjects of study to be discussed in Chapter Five. On condition that the sample size is adequate, this data enables significant subgroups to be identified when discussing the variables influencing international students' perceptions of the IELTS test.

Part 3 was designed to collect data on test results and to identify respondents who had taken both the IELTS and the TOEFL. The number of times students take a language proficiency test may influence their perceptions of the test. According to Gibson and Rusek (1992), IELTS (Australia) suggested that an overall band score of 6.0 is generally acceptable for entry to tertiary level studies. This is considered equivalent to a score of 550 on the TOEFL. The questionnaire included test result categories below Band 6 to take account of respondents following an English language course and attempting to improve their IELTS test score for entry to academic studies.

Calculating a mean IELTS score from four subtests may result in a candidate's overall band score falling between the single gradations from Band 1 to Band 9 of language ability on the proficiency rating scale (see Appendix II). This is because the Reading and Listening subtests are objectively rated with 0.5 gradations between Band 1 and Band 9, whereas the Writing and Speaking subtests are subjectively rated with single gradations. The issue then is whether a band score of 5.5, for example, is closer to the descriptor of Modest User, Band 5, or Competent User, Band 6. This is problematic since proficiency rating scales do not provide well-defined cut-off points between band scores, as discussed in Chapter Two. The researcher was aware that some departments in the university were accepting students as undergraduates in certain circumstances with a band score of 5.5. Furthermore, on completion of an ELICOS course, students with an IELTS band score of 5.5 might be

accepted for tertiary studies by colleges in Australia. The researcher therefore chose a subjective categorization of IELTS band scores to include 5.5 and 6.0 in the same category on the questionnaire (see Appendix VI). Since a score of 550 on the TOEFL test is the usual minimum entry requirement to Australian universities, the categorization of TOEFL scores on the questionnaire was based on this cut-off point, and on the TOEFL score range groups used by Light *et al.* (1987) in their study of the relationship between TOEFL scores and academic success.

Part 4 of the questionnaire was designed to gather data on perceptions of test content. Do international students perceive the IELTS to be a test of English for Specific Purposes or a test of general language proficiency? Do the test takers think that grammar and vocabulary are being tested in addition to the four macroskills? This section also obtained information about the modules chosen by students and assessed the influence of background knowledge on test performance, as discussed in Chapter Three.

Part 5 of the questionnaire firstly ascertained students' perceptions of the recommended minimum requirement of Band 6 for tertiary studies. Is this cut-off point realistic for academic studies? Individual universities in Australia have set their own minimum English language proficiency entry requirements, but these have generally followed the IELTS (Australia) guidelines. Higher entrance levels may be required for postgraduate studies, or in particular departments and faculties, and minimum subtest band scores may be required at some universities. In contrast, the British Council recommends a minimum overall band score of 6.5 to 7.5 for academic studies in the United Kingdom (IELTS, 1992: 5). Secondly, Part 5 of the questionnaire determined the relationship between students' perceptions of their test performance and their language ability. This section also aimed to find out if the test takers think the IELTS test predicts a student's readiness for academic study in English, or if they perceive other factors to be influential.

Part 6 of the questionnaire asked students for their views on the 'speededness' of each subtest. It also examined test takers' perceptions of the diagnostic profile provided by the IELTS test. Respondents were asked to rank their test performance on each macroskill and to judge the difficulty level of each subtest. Part 7 of the questionnaire aimed to assess students' perceptions of the importance of each macroskill in the target situation and the authenticity of reading and writing tasks on the IELTS test.

Interview

There is some debate over the extent to which an interviewer should explain the meaning of questions to respondents, or should correct misunderstandings. Fowler and Mangione (1990) argued that respondent confusion indicates poorly worded questions, and that further interpretation by the interviewer reduces the reliability of the results because respondents then effectively answer different questions. They recommended the repetition of questions, and use of the terms 'how do you mean that?', 'tell me more about that' and 'anything else?' as non-directive probes of interviewees (Fowler and Mangione, 1990: 42). However, they also cited Mishler (1986) who argued that there should be scope for interviewer initiative and interviewee narrative. The interview design in this study (see Appendix VII) followed a 'semistructured approach involving the asking of structured questions followed by clarifying unstructured, or open-ended questions' (Gay, 1992: 232).

There were four main objectives of the interview. First, the interviewees were asked to complete the background data section to provide a profile of the subjects of the group interview. Second, the needs of the target situation were assessed by identifying the language skills required for academic study. The third objective was to establish whether the group shared an understanding of the construct of language proficiency. The fourth objective was to compare students' perceptions of the nature of two standardized language proficiency tests, the IELTS and the TOEFL, particularly regarding the test methods used.

Procedure

The first task was to distribute questionnaires to all international students. The second task was to identify those respondents who had taken an IELTS test. The third task was to follow-up the questionnaire by interviewing a sample of respondents. The final task was to analyse the data collected. The researcher was fortunate in obtaining the endorsement of this study by the International Student Office.

As information concerning test results is confidential, it was not possible to send questionnaires only to those students who had taken the IELTS test. Questionnaires were mailed to all international students enrolled at the Launceston campus. 240 questionnaires were distributed in total, excluding those returned because of students changing their address, moving to the Hobart campus of the university, or returning overseas. It was hoped that endorsement by the International Student Office, where the survey results would be available on completion of the research, would encourage the target population to respond.

An accompanying letter, two statements of consent and an information sheet were enclosed with the questionnaire (see Appendices IV and V). International students were asked to return one copy of the statement of consent and the questionnaire by a fixed date, which allowed two weeks for responding, excluding mailing time. A survey box was placed outside the International Student Office on campus so that respondents need not be involved in any expense when returning the questionnaire. The response rate to the survey was approximately 20% by the deadline date. A personally signed reminder letter to all non respondents increased this figure to 32% (78 respondents) after a second deadline two weeks later. No partially completed questionnaires were received which would render them invalid. Thirty-eight respondents had not taken an IELTS test; therefore, the forty respondents who had taken the test became the final sample for this research.

One objective of the questionnaire was to identify respondents who had taken both the TOEFL and the IELTS tests, and to ask those respondents if they were willing to be interviewed by the researcher. In fact, only thirteen respondents had taken both standardized tests. A sample of five respondents was chosen for group interview, not including students interviewed in the pilot study, on the basis of their availability for interview. At the time of the group interview, each interviewee was given a copy of the interview format as a guideline. It was decided to tape-record the interview to ensure an accurate and precise record for analysis. A copy of the interview tapescript is available in Appendix VIII.

Limitations of Methodology

A small-scale study of forty international students does not allow generalizations to the target population, that is, to all international students studying in Australia who have taken an IELTS test. The students surveyed at the Launceston campus were by definition not a representative sample of all international students in terms of nationality and academic studies. Due to the confidentiality of test results and no central database of IELTS test records, the researcher is unable to assess how representative the sample is of all international students on the Launceston campus who had taken the IELTS test before April 1994. Some international students enter academic courses on the basis of a TOEFL score, and others with high school qualifications after completing senior secondary education in English in their home countries.

In addition, the study was dependent on using volunteers amongst the international student population, with no obligation to participate in the survey; this raises the likelihood of sample bias. Some students might have felt inhibited by their language ability from responding to the questionnaire. Furthermore, 'respondents from some cultural groups may be more likely than others to report what they think the researcher wants to hear' (Turner, 1993: 737). The study

did not sample international students who had been refused admission to an academic course at an Australian university because of poor performance on the IELTS test, apart from a small number studying in the English Language Centre. Their perceptions of the test might be quite different from students who have achieved the minimum entry requirement.

Were the students able to access the information the researcher required on the questionnaire? Certainly questions about past events create problems of reliability of recall. One difficulty in the design of the questionnaire was how to help respondents access information about a past event, especially if the test-taking experience was a negative one. Some form of prompt might have been helpful. However, it is the researcher's opinion that international students probably have a high level of recall about a language proficiency test which was very influential in their gaining admission to an Australian university.

In some parts of the questionnaire, the non-substantive filter 'not sure' was omitted and the researcher's concepts were imposed upon the respondents. The lack of a non-substantive filter, in some cases, was due to the researcher's belief that cultural response factors might result in overuse of a 'not sure' category. However, by using closed questions 'we do impose either our own view of reality, or our guesses about our respondents' views of reality, upon our respondents' answers' (Foddy, 1993: 192).

A number of studies cited in Fowler and Mangione (1990) have pointed out the innumerable pitfalls of interviewing strategies, including the tendency for directive rather than non-directive probing of interviewees who are non-native speakers of English, and the difficulty of retaining a neutral, nonjudgemental attitude to respondents' subjective comments. Nevertheless, a focused group interview which depends on the 'prior analysis by the researcher of the situation in which subjects have been involved' has advantages in terms of interview design (Cohen and Manion, 1980: 259). The format of the interview encouraged unanticipated responses, the sharing of subjective experiences, and the clarification of

language testing concepts which contributed to interpretation of data gathered from the questionnaire survey.

The low response rate to the questionnaire limits generalizations that can be made about international students' perceptions of the IELTS test. Those who responded may feel more strongly about the test than non respondents for a number of reasons. 'For questions about subjective phenomena, such as opinions, feelings, or perceptions, there does not exist even in theory a direct way to assess the accuracy of an answer.' (Fowler and Mangione, 1990: 24). Furthermore, nominal scales to gather categorical data were chosen for the questionnaire design and it was recognized that this choice would limit the methods of statistical analysis possible.

Summary

An investigation into international students' perceptions of the IELTS test used a survey questionnaire and focus group interview. Questions focused on perceptions of test content, Band 6 proficiency level, the IELTS test as a predictor of academic success, test time, level of difficulty, the target situation, task authenticity and relevance of background knowledge. In addition, the interview also focused on comparisons of test methods between the IELTS and the TOEFL tests.

Despite the small-scale nature of this investigation, considerable data was generated from the questionnaires. Many respondents submitted copious additional comments. Moreover, the focus group interview provided an insight into how students view language proficiency tests. The data is analysed and interpreted in Chapter Five. This study may provide a starting point for establishing a methodology for assessing international students' perceptions of second language proficiency tests.

Chapter 5

Data Analysis and Interpretation

Introduction

A sample of forty subjects provides the response data set from the questionnaire survey. All subjects fulfilled two conditions for inclusion in the sample: they had taken an IELTS test before April 1994; and they were studying on the Launceston campus of the University of Tasmania, semester I (1994). No figures were available for the total number of international students at Launceston who had taken an IELTS test. However, for entrance to Australian universities international students can also be selected on the basis of a TOEFL test score, or with high school qualifications obtained through English-medium secondary education. The data response sample ($n=40$) included all those international students following academic courses on the Launceston campus who had completed an intensive English language course in the English Language Centre between 1992 and March 1994. For this reason, although a sample of forty is small, it may represent a significant proportion of the target population of IELTS test takers at the University of Tasmania, Launceston.

Sample Characteristics

More than 92% of all international students studying at Launceston are Asian. Data from the questionnaires shows that 95% of the subjects in the research sample were international students from Asia, primarily Malaysia, Thailand and Indonesia (see Figure 5.1). Therefore, the first languages of the sample were also predominantly Asian. However, subgroups of nationality and first languages were too small for analysis in this study. 55% of the sample

were males and 45% were females. The age range of students was from 18 to 37 years, with an average age of 24.8 years. In summary, the sample of forty international students in this study represented a diverse group of Asian students, with approximately equal numbers of males and females, 50% of whom were between 18 and 22 years of age.

Figure 5.1
Nationality of subjects (n=40)

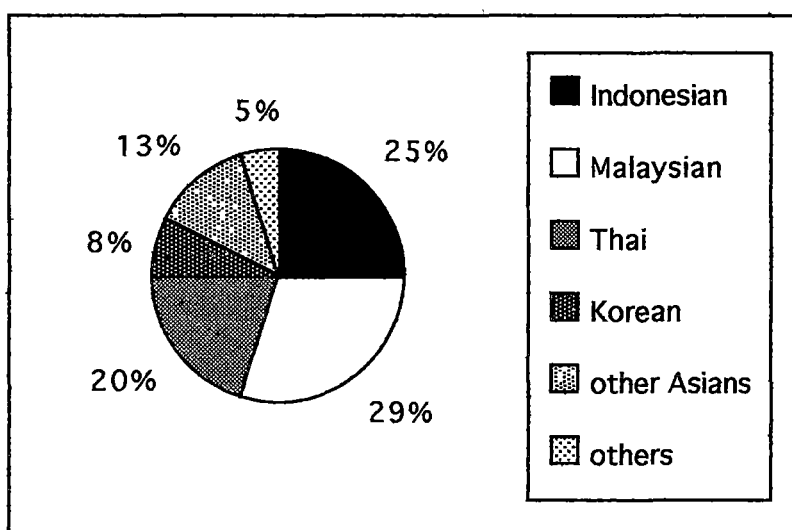


Figure 5.2 shows the period of time that the subjects of this study had been in Australia. 42% of respondents were in their first year of study. In addition, 57.5% of respondents were undergraduates, 27.5% were postgraduates and 15% were ELICOS students. Table 5.1 shows the range of academic courses studied by the sample of international students. 87.5% of students (35) in the survey had studied in English language classes in Australia. 20% of respondents (8) had never taken English language classes in their own country, whereas 45% (18) had studied English for more than six years in their own country.

Figure 5.2
Years in Australia (n=40)

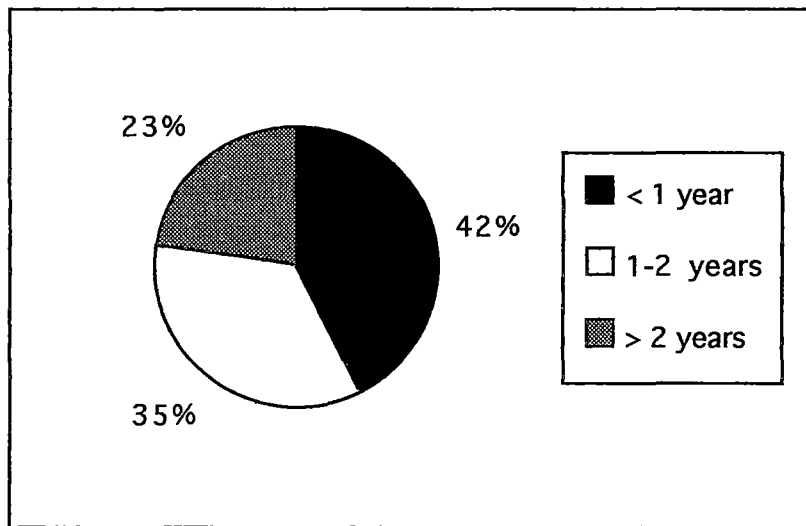
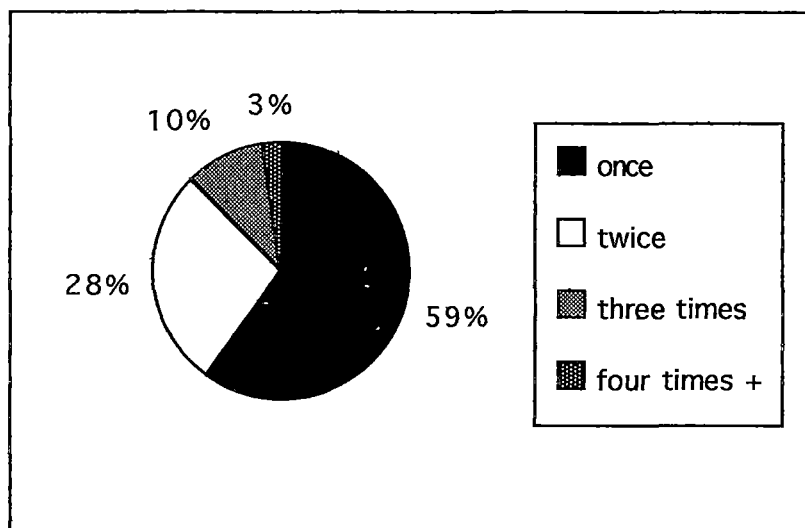


Table 5.1
Course of study (n=40)

Course of study	N
Computing	8
English	6
Aquaculture	6
Business	5
Environmental Design	5
Education	4
Architecture	3
Medical Lab. Science	3

The number of times test takers sit for a language proficiency test may influence their attitudes towards the test. Figure 5.3 shows that 59% of the respondents (24) had taken the IELTS test once, and 28% (11) had taken the test twice. Therefore, most of the subjects had limited experience of taking the IELTS test. 70% of the international students in the sample had sat for an IELTS test in their own country and 42.5% had sat for the test in Australia. Only eight respondents had taken the IELTS test more than two years prior to March 1994 when the questionnaire survey was undertaken. To summarize, 80% of the sample population had taken an IELTS test within the last two years.

Figure 5.3
Number of times IELTS test taken (n=40)



87.5% of the research sample (35) had an overall band score of 5.5 and above on the IELTS test. Although Band 6 is generally the minimum requirement for entry to an Australian university to pursue an academic course, international students are sometimes accepted for undergraduate courses with a band score of 5.5 in cases of academic potential. As Table 5.2 illustrates, only five respondents had a score below 5.5 on the IELTS test, four of whom were following a course of instruction in the English Language Centre. In other words, they

were aiming to improve their IELTS band score or their general language proficiency level. 32.5% of respondents (13) had taken both an IELTS test and a TOEFL test.

Table 5.2
IELTS overall band scores (n=40)

IELTS Band Score	N
< 4.0	1
4.5-5.0	4
5.5-6.0	17
6.5-7.0	15
> 7.0	3

Test Content

Table 5.3 shows respondents' perceptions of the content of the IELTS test (n=40). 75% of respondents (30) thought that the IELTS tested a lot or some grammar accuracy, although grammar does not appear overtly as a separate subtest in the IELTS. Furthermore, most respondents thought that both general vocabulary and subject-specific vocabulary were significant in the test, with 47.5% (19) judging that a lot of general vocabulary was tested, and 40% (18) judging that a lot of subject-specific vocabulary was tested. Most respondents perceived the Listening subtest to be general rather than subject-specific (see Table 5.3). 62.5% of respondents (25) thought that there was a lot of general listening, and 20% (8) thought that there was a lot of subject-specific listening in the test. In contrast, the Writing subtest was generally judged to be subject-specific. 50% of respondents (20) thought that

there was a lot of subject-specific writing in the test, whereas only 25% (10) thought that there was a lot of general writing.

Table 5.3
Perceptions of test content
(g= general) (s-s = subject-specific)

	a lot	% N	some	a little	none
Grammar accuracy	5	(12.5)	25	9	1
Vocabulary (g)	19	(47.5)	18	3	0
Vocabulary (s-s)	16	(40)	19	5	0
Reading (g)	17	(42.5)	20	2	1
Reading (s-s)	14	(35)	22	4	0
Writing (g)	10	(25)	23	6	1
Writing (s-s)	20	(50)	18	2	0
Listening (g)	25	(62.5)	12	3	0
Listening (s-s)	8	(20)	19	12	1
Speaking	14	(35)	23	3	0

However, the distinction between ESP and general proficiency testing was less clear on the Reading subtest. 42.5% of respondents (17) perceived a lot of general reading in the test, and 35% (14) perceived a lot of subject-specific reading in the test. This may reflect a limitation in the design of the questionnaire as the criteria for judging whether content is general or subject-specific was not defined for the respondents. Background knowledge applicable to the reading comprehension passages might have been categorized as general or subject-specific by respondents according to their interpretation of the question.

Students were not asked to judge the Speaking subtest as general or subject-specific since the oral interview, although structured, varies in content according to interviewer and interviewee. The Speaking subtest represents the smallest amount of testing time given to any of the four macroskills, and yet 35% of respondents (14) thought that a lot of speaking was tested in the IELTS (see Table 5.3).

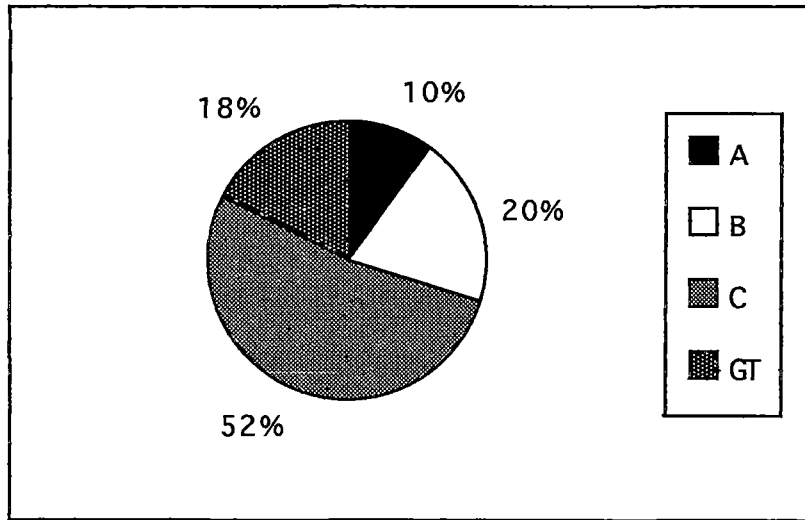
Some comments on test content by respondents suggested that the IELTS encourages test preparation and proficiency in all language skills:

- IELTS tests all four skills which is good.
- By taking IELTS test I improved and balanced my skills in reading, listening, speaking and writing in English.
- I think IELTS test has equal parts of grammar, vocabulary, reading, writing, listening; only the speaking part seems to be the least.

In summary, international students perceived that the IELTS (1989 version) tested grammar although it did not appear as an independent subtest. Furthermore, lexis was considered to be significant, both general and specialized. The Listening subtest tended to be perceived as a general academic test, whereas the Writing subtest was generally recognized to be ESP. However, respondents' perceptions of reading as either subject-specific or general may have depended on the amount of background knowledge in a specific subject that they were able to apply in the Reading subtest. Perceptions seemed to be unrelated to the ESP module chosen for the test.

As stated in Chapter Two, 75% of international students who take the IELTS test in Australia choose module C for the ESP subtests (see Appendix III). Figure 5.4 shows that only 52% of the research sample selected module C. However, 57.5% of the respondents (23) had never taken an IELTS test in Australia.

Figure 5.4
Modules chosen by students (n=40)



26 respondents stated they had background knowledge in a specific subject for the ESP module chosen for the IELTS test. Of these respondents, 74% thought that background knowledge was very helpful or helpful in the test. Additional comments suggested that some respondents perceived the broad fields of academic study included in each module to preclude their use of subject-specific background knowledge in the test:

- I think it's not fair because each module covers too wide subjects.
- Sometimes it is not quite helpful. For example, in module B the test can be medical science or botany or zoology sciences. In module A also can contain many technical sciences.

In contrast, some students were able to use their background knowledge in the test:

- This is one primary advantage of the IELTS test.
- I used a lot of my background knowledge to complete the writing sections.
- The subject matter may lead me to find the answers easier, even though I don't understand all the reading.

The significance of these findings will be summarized in Chapter Six when considering the implications of a revised 1995 version of the IELTS test.

Proficiency Level

Band 6 on the IELTS test is the recommended minimum requirement for entry to academic courses at the tertiary level in Australia. Table 5.4 shows that 57.5% of respondents (23) perceived Band 6 to be a good level of English and 32.5% (13) perceived Band 6 to be a fair level. On the IELTS test proficiency rating scale, Band 6 is rated as Competent User and Band 7 as Good User (see Appendix II). The hypothesis that there is no significant relationship between respondents' perceptions of Band 6 and their level of academic study was tested using the Chi-square test. The null hypothesis was upheld, but the small sample size limited any conclusive interpretation as only eleven of the respondents were postgraduates.

Table 5.4
Perceptions of Band 6 by level of study (n=40)

	undergraduate	postgraduate	ELICOS	% N
very good	1	0	2	7.5
good	14	5	4	57.5
fair	7	6	0	32.5
limited	1	0	0	2.5

In addition, no significant relationship was found between respondents' ratings of Band 6 and their individual IELTS test band score results. The results of a Chi-Square test are

therefore not presented here. In other words, within the limitations of the small sample size, international students' judgements of Band 6 seemed to be unrelated to their language proficiency test results: 90% of students (36) thought that an overall IELTS test result of Band 6 was good or fair (see Table 5.4).

Some respondents' comments suggested that Band 6 is only realistic as a baseline or minimum entry requirement, and several students noted that it depends on the particular course of study:

- Although I could get pass the score which was required but it's not good enough to study in the university.
- For me that score is not standard to be successful in Australian university. But it is a very good basic to improve our English ability.

90% of respondents thought that Band 6 is a realistic entry level for an undergraduate course at a university in Australia. However, although 35% thought that Band 6 is realistic for a postgraduate course, 52.5% thought that it is too low. A summary of comments supports these findings:

- There is still a lot of problems but you may be able to cope if you study hard.
- For undergraduate more than 6.0 is better, but 6.0 is okay.
- For postgraduate it should be slightly higher, say 6.5 or 7.0.
- Postgraduate course requires higher skill in writing, especially for academic use, as well as in speaking in order to communicate and exchange ideas.

Table 5.5 shows students' perceptions of their individual overall performance on the IELTS test rated on a scale from limited to excellent. The data is tabulated according to students' IELTS test scores gathered in Part 3 of the survey questionnaire (see Appendix VI). There was a possible relationship between students' perceptions of their performance on the IELTS

test and their band score results. A Chi-Square test showed a less than 0.05 probability level that the relationship was due to chance (see Table 5.5).

Table 5.5
Perceptions of test performance (n=40)

band score	<4.0	4.5-5.0	5.5-6.0	6.5-7.0	>7.0
very good	0	0	2	6	2
good	0	1	3	8	1
fair	1	2	11	1	0
limited	0	1	1	0	0

DF	12
Total Chi-Square	23.095 (p =.0269)

Table 5.5 shows that most students who scored 5.5 or 6.0 on the IELTS test rated their individual overall test performance as fair. Most students who scored 6.5 or 7.0 rated their performance as good or very good. Ratings of test performance seemed to match the IELTS test proficiency rating scale descriptors of Competent User (6.0) and Good User (7.0).

Table 5.6 shows respondents' perceptions of their English language level when they took the IELTS test. The data is tabulated according to students' IELTS test scores. There was a probable relationship between students' perceptions of their English language level, on a rating scale from limited to excellent, and individual overall band score results on the test. A Chi-Square test showed a less than 0.05 probability level that the relationship was due to chance (see Table 5.6). Only 14 respondents rated their level of English as good when they took the IELTS test, and 16 rated their level as fair.

Table 5.6
Perceptions of English language level (n=40)

band score	<4.0	4.5-5.0	5.5-6.0	6.5-7.0	>7.0
very good	0	0	0	3	0
good	0	0	6	5	3
fair	1	1	7	7	0
limited	0	3	4	0	0

DF	12
Total Chi-Square	24.117 (p =.0196)

In short, a closer examination of students' perceptions of overall performance on the IELTS test and perceptions of English language level at the time of taking the test revealed that students rated their language level lower than they rated their overall test performance on the rating scale from limited to excellent. In other words, it seemed that most students performed better on the IELTS test than they expected. This may also be due to Asian international students tending to underrate their individual level of test performance. Some respondents were evidently disappointed with their performance:

- IELTS test does not fully reflect my performance and level of English since it only shows basic skill.
- As far as I'm concerned there is a time limit putting us under pressure. As a result we might not perform well.
- I only managed to do well in certain parts and fairly well in others.

In addition, comments from some students suggested that perceptions of English language ability change once they enter academic courses in Australia:

- We will feel our level is very good comparing to other foreign students who come from country where English is their second language. However, it will be very different if we compare our ability to native speakers or people who come from the country that use English intensively.
- My English level improved when I started studying in Australia.
- I had never taken any English course except English class in high school when I took the IELTS test. I never use English except in reading books. My English therefore was only a passive ability and I only relied on memory and this passive skill in taking my first and only test.
- People can improve when experiencing something because not many have chances to practice the English language in their own country.

The IELTS Test as a Predictor

80% of respondents thought that the IELTS test predicts an international student's ability to study in English. However, question 5, Part 5 (see Appendix VI) revealed a limitation in questionnaire design as it did not ask how 'accurately' the IELTS test predicts ability to study in English. Fortunately, students added many comments to this section of the questionnaire which are summarized in the following statements:

- Because the result of IELTS is very fair and objective, the ability of students of English skills can be seen from that test as real scores.
- Mainly for postgraduate studies, much better than TOEFL.
- Yes, especially if the test material really similar with the individual background knowledge.

Although 80% of respondents thought that the IELTS test predicts ability to study in English, their comments suggested reservations:

- IELTS indicates ability in general. However, different levels and different portions of skill are needed in studying different fields.
- It does a little, but the time allocated for the test is not enough to predict a student's overall

ability to study in English. Taking a test is different from listening and taking notes from lectures.

- Yes, as far as the candidate has a good ability to manage time.
- Partly it does but mostly people are improving with time.
- It is only one method to be the criteria of English measurement.
- However, it is difficult to predict someone's ability just from one test. For example, band score 5.0-6.5 it seems there is no significant difference. It is likely IELTS may describe someone's ability if the test is fit exactly with student's background knowledge.

Clearly one respondent felt that the IELTS test does not predict ability to study in English:

- No because one test cannot predict much, especially when students are panicked in test.

36 respondents (n=40) thought that Band 6 is a realistic entry level for undergraduate study. No significant relationship was found between level of academic study and respondents' perceptions of Band 6 as a realistic entry requirement for an undergraduate course. The results of a Chi-Square test are therefore not presented here. However, there was a probable relationship between level of academic study and respondents' perceptions of Band 6 as a realistic entry requirement for postgraduate study (see Table 5.7). A Chi-Square test showed a less than 0.05 probability level that the relationship was due to chance. 21 respondents thought that Band 6 is too low for postgraduate study, and 14 thought that it is realistic.

Table 5.7
Perceptions of Band 6 as realistic (n=36)
(postgraduate study)

	undergraduate	postgraduate	ELICOS
too high	0	0	1
realistic	4	7	3
too low	15	4	2

DF	4
Total Chi-Square	11.414 (p =.0223)

Other major factors identified by international students as predicting ability to study in English, apart from the IELTS test were: first, academic background; second, hard work and motivation; third, country of origin; and fourth, verbal participation in the classroom and social interaction outside the classroom. The significance of country of origin is shown by the following respondents' comments:

- Students from country where English is used equally with their native language have higher ability to study in English than students from other countries where national language and dialects are mainly spoken.
- If English is equally used with native language, like in Singapore and Malaysia, students from these countries will have high ability to study in English. This is obviously different from students from other countries where English is not used as every day formal communication like Indonesia.

Test Time

Table 5.8 shows that most respondents thought that adequate time was allowed in the IELTS test for the subtests of Listening and Speaking. In contrast, only 52.5% of students (21) thought that the time given for the Writing subtest was adequate, and 35% of students (14) thought that too little time was given for the Reading subtest. Only one respondent suggested that there was too much time given to the Reading subtest. There were few comments concerning the 'speededness' of the test. In summary, the ESP components of the IELTS test were generally perceived to be constrained by time compared with the general proficiency subtests. 'By introducing an element of speededness, we are discriminating among persons

not only in their knowledge of the facts, but in the speed with which they are able to report that knowledge.’ (Henning, 1987: 79).

Table 5.8
Perceptions of test time (n=40)

time	adequate	too little
reading	25	14
writing	21	19
listening	32	8
speaking	33	7

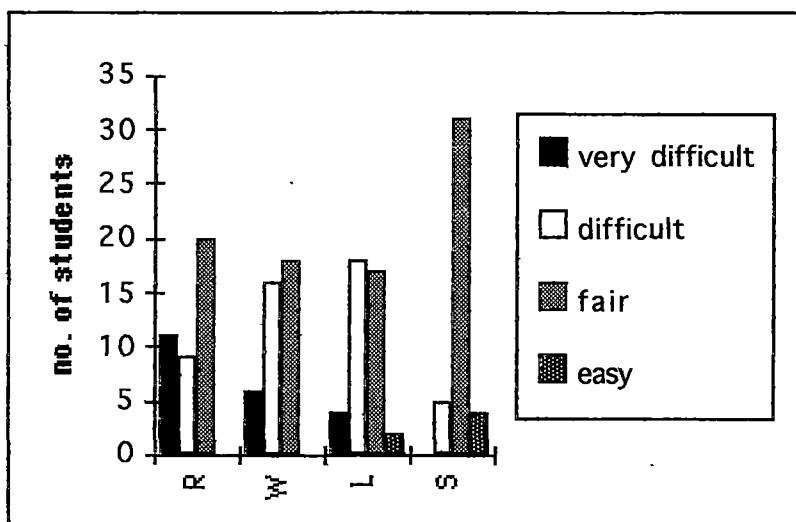
Diagnostic Profile

Figure 5.5 shows frequency distributions for perceptions of level of difficulty of the IELTS subtests on a scale from easy to very difficult. 55% of respondents (22) thought that the Writing subtest was difficult or very difficult, 52.5% (21) thought that the Listening subtest was difficult or very difficult, and 50% (20) thought the Reading subtest was difficult or very difficult. The Reading subtest was judged to be very difficult by 11 respondents. On the other hand, 50% of respondents (18) thought that this subtest was fair. 77.5% of respondents (31) thought that the Speaking test was fair and no respondents rated it as very difficult.

To summarize, responses tended to point to similar perceptions of difficulty for both the subtests of Writing (W) and Listening (L) as shown in Figure 5.5. The Speaking subtest (S) was perceived to be fair by most students, but responses to the Reading subtest (R) were

diverse. Responses appeared to be unrelated to either the module chosen or to perceived background knowledge in a specific subject. There were no comments from respondents adding further explanation.

Figure 5.5
Perceptions of level of difficulty of subtests (n=40)



However, students' ratings of perceived level of difficulty of each subtest appeared to be related to individual rankings of some IELTS subtest results. Respondents were asked to rank their individual subtest results on a scale from highest to lowest. There seemed to be a significant relationship between respondents' ratings of the level of difficulty of the Reading subtest and the rank order of Reading subtest results (see Table 5.9). A Chi-Square test showed a less than 0.01 probability level that this relationship was due to chance. Furthermore, there was a probable relationship between respondents' ratings of the level of difficulty of the Listening subtest and the rank order of Listening subtest results. A Chi-Square test showed a less than 0.05 probability level that this relationship was due to chance (see Table 5.10).

Table 5.9

Rank result and level of difficulty (Reading)

DF	6
Total Chi-Square	18.215 (p =.0057)

Table 5.10

Rank result and level of difficulty (Listening)

DF	9
Total Chi-Square	18.251 (p =.0324)

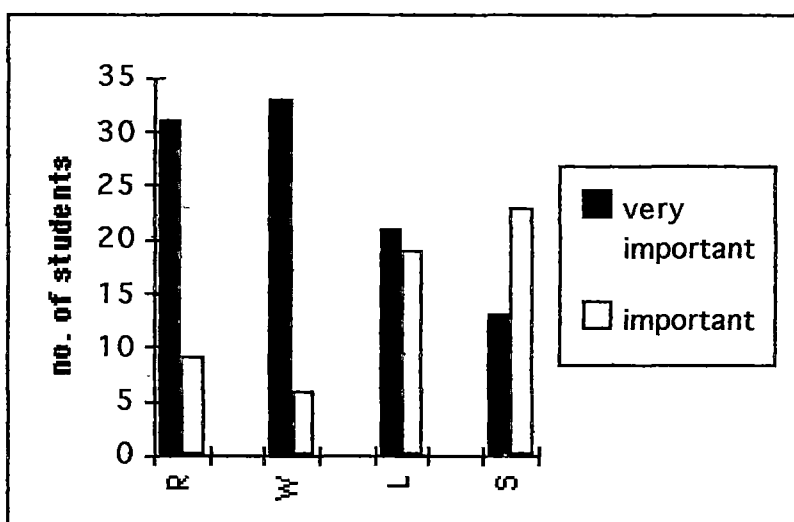
There was no significant relationship found between the ratings of perceived level of difficulty and rank order of Writing subtest results, nor between ratings of level of difficulty and rank order of Speaking subtest results. In other words, performance on the Reading and Listening subtests seemed to be related to respondents' perceptions of levels of difficulty of each subtest. However, performance on the Writing and Speaking subtests seemed to be unrelated to perceptions of levels of difficulty. These findings cast doubt on the appropriateness of a Chi-Square test for statistical analysis of the small sample size (n=40) in this survey. In addition, individual rankings of IELTS test results are not as reliable as the actual diagnostic profile of band score results.

Target Situation

As Figure 5.6 shows, international students considered reading (R) and writing (W) to be the most important skills required for academic study. 31 respondents (n=40) rated reading as very important, and all other respondents rated these skills as important in the target situation. 33 respondents rated writing as very important, and 6 rated writing as important. Figure 5.6 shows that 21 respondents rated listening (L) as very important and all other respondents rated this skill as important for academic study. Speaking (S) was rated as very important by 13 respondents, and as important by 23 respondents.

Figure 5.6

Importance of macroskills in target situation (n=40)



Students' comments suggested they perceived all skills to be important in the target situation; nevertheless, particular courses of study might attach more importance to reading and writing skills:

- Four of these aspects are interconnected.
- All parts in IELTS are needed and important for study.

- Everything is very important because English language is widely used in the real life situation.
- In my opinion, to study well in whatever course everyone should take each skill importantly.
- All aspects are needed by students, but writing and reading are the most important one to support study for postgraduate level.
- As my study is by research, reading and writing skills are more required than listening. Speaking is also important as I have to do presentations, seminars as well as discussions with relevant experts.

Authenticity

72.5% of respondents (29) thought that the writing tasks in the IELTS test were similar to those given in their course of academic study. 62.5% of respondents (25) thought that the reading tasks in the IELTS test were similar to those given in their course of study. Students' comments generally suggested that tasks in the target situation are considerably more demanding than in the IELTS test:

- In real course it's much more difficult.
- In my course of study greater skill and different style in writing are required.
- Basically, how to write an essay, an idea or an assignment are the main purpose of that test skill.
- Similar in the sense that reading, understanding and interpreting in writing tests one's ability just as does the IELTS test.
- Reading and writing in IELTS is the basic knowledge (skills) for me to do any assignments/dissertation.
- IELTS test is very up to date system test English as the questions on it always different from each to another period test. We could not use memorize skills to deal with the test, but we should have reflex skills. This is very relevant to academic culture.

Subjects of Interview

According to data on the questionnaire survey, thirteen respondents had taken both the IELTS and the TOEFL tests. Two of the respondents took part in the pilot interview and two were unwilling to be interviewed. Eight of the thirteen respondents came from Indonesia and ten were studying at postgraduate level. The five subjects chosen for the focus group interview (May 1994) had taken the IELTS test most recently. Four of the international students interviewed had taken the IELTS test within the last year. The focus group consisted of three Indonesians, one Thai and one Japanese student. Four were postgraduates and one was an ELICOS student planning to undertake postgraduate studies. A questionnaire format (see Appendix VII) was used in the focus group interview, and a full transcript of the interview is shown in Appendix VIII.

Target Situation of Interviewees

The focus group confirmed the importance of all four macroskills, reading, writing, listening and speaking, in the target situation of academic studies at the University of Tasmania. However, it was recognized that the significance placed upon each macroskill depends upon level of study as well as academic discipline:

- S2 Sometimes for academic purpose I think speaking is not important. Especially for masters degree with research the speaker is the supervisor. Maybe no need to go to classroom to discuss, so it depends.

(R- researcher; S1- student one; S2- student two; S3- student three; S4- student four; S5- student five).

The Nature of Language Proficiency

A consensus was reached on the nature of language proficiency amongst the group. The group defined language proficiency as readiness for academic study in English:

- S3 The ability to perform in English.
- S4 I think English proficiency here means our ability to do courses in universities, including reading, writing, and speaking, listening, all things especially in academic aspects.
- R Do you see proficiency as something divided into skills: reading, writing?
- S1 No
- S5 Includes everything.
- S1 Overall.

Background Knowledge

The focus group interviewees differed in their perceptions of the IELTS as a test of English for Academic Purposes. The issues of test content and background knowledge were discussed by the group without a consensus being reached. Some students felt that the IELTS test did not reflect their English language ability as their background knowledge was not relevant to the content of the ESP subtests. Other students perceived the Reading subtest to be more general in nature. The integration of reading comprehension texts and writing tasks seemed to enhance the face validity of the IELTS test:

- S1 I mean anybody they don't understand the field can be guided by the test because of their previous reading, because of their reading at the previous level. Same field there are several readings before you proceed with the writing in special field. You are given the reading material, the reading regarding that field.

- R What do you mean 'given it', given it by whom?
- S1 Given it by the exam.
- R You mean in the exam paper you have the reading passages?
- S1 Yes. I think it guides you.

In contrast, the group perceived the TOEFL test to be a general proficiency test with less EAP emphasis and relevant to a wider target situation:

- S1 It seems that TOEFL tests your comprehensive knowledge not academic.
- R What do you mean comprehensive knowledge?
- S1 Your vocabulary, your grammar.
- S2 I'm not sure. Probably I agree with ... (S1). TOEFL tests for studying. This test also applicable for someone who just go abroad. This test only assess English, but not academic purpose.
- R So your background knowledge, is that helpful in the TOEFL test?
- S4 It is general.

S4 explained this point further when comparing the IELTS test with the TOEFL test :

- S4 The first one I think we have discussed. IELTS test more disposed to the academic purpose rather than the TOEFL, it's general.

Further Comparisons

The group focused on objective scoring in the design of the TOEFL test compared with the IELTS test. This issue was not addressed in the questionnaire survey. They concluded that a speaking test is problematic in the TOEFL test if objectivity of scoring is the aim. Another point of comparison highlighted by the focus group was the inclusion of a grammar

component in the TOEFL test, whereas grammar is not tested separately in the IELTS. A further point raised was the norm-referencing of results on the TOEFL test compared with a perception that the IELTS test proficiency band scores are 'independent':

- S5 Yes, I think TOEFL has grammar section. IELTS test does not have grammar section so it is really different. Also TOEFL each question is independent, but IELTS test dependent question. It is different I think. IELTS is severe, but TOEFL not severe I think.
- S3 One difference is that TOEFL test depends on the student, the people who test. The score, it depends on how high the average score or something like that.
- S5 IELTS independent.
- S3 Could you understand me? The score depend on the group.

Test Methods

Interviewees were clearly aware of test method effects on test performance. The strategy required to succeed in the TOEFL test is apparently not transferable to the IELTS test. As indicated in the next extract, the IELTS test was perceived to favour test methods other than multiple-choice items. Time management was also mentioned, although time is obviously a constraint in most test settings. The factors mentioned related to the test taker's prior experience of a test. Prior experience and training develop strategies for efficient test taking which enhance 'test wiseness' (Bachman, 1990: 114):

- S2 I think first time really need an English course for a couple of months. Of course it depends on the background, I mean the ability of the student because student who has English ability poor maybe they need longer in English class than for student has better English background. Then, of course, we still need to read past tests because in my experience I got 550

TOEFL, then first time I try to fill the exercise IELTS test I got lower mark than other students who has already experienced in filling the IELTS test. Then after a couple of weeks I find myself much better than them because I think my background English is already better. Then I only need the skill how to fill this test.

R So you mean there's a skill, a strategy for the test?

S2 Right.

S1 I think interaction, by simulating interaction with others, reading and discussion. Independent study of course more and more reading, and writing skill.

R Anybody else?

S3 One problem to prepare the student for the test is to give the strategy to manage the time to take the test.

As the next extract indicates, the group proposed that multiple-choice practice is necessary for the TOEFL test, and they felt that this test method might encourage guessing. Bachman (1990: 128) suggested that when test takers are under pressure because of the perceived 'speededness' of a test, guessing or other compensatory strategies may be used by the test takers. The evidence indicated that this type of response strategy occurred:

S4 In the TOEFL test strategy also is very important, especially to manage time because it's very hard, and the reason why, if I compare with IELTS test, is in TOEFL we should interpret what the sentence means because there are four many choices and almost similar, so it's very hard, but in the IELTS test we just fill in the gaps with phrases so we need skill also for TOEFL actually.

S3 For guessing.

S4 Guessing, if you want to guess. Yes, also we can guess. We can get a good mark on this guessing, but for IELTS test if you don't know it you cannot do.

The emphasis on grammar in the TOEFL test and the prominence of multiple-choice as a test method was highlighted once more by interviewees:

- S5 Grammar section is very easy to get high score in TOEFL, if you do exercises, lots of exercises.
- R Lots of practice.
- S5 Lots of practice for grammar is very easy to get higher score.
- R So from what you've said you can see different test methods used between the IELTS and the TOEFL. So if you just wanted to summarize, what would you say the main differences in the test methods used are between IELTS and TOEFL? but sometimes we have multiple-choice, but hardly I think. The TOEFL all are multiple-choice The problem is that some student can guess.

Authenticity of Tasks

Bachman stated that 'the absence of interaction that characterizes adaptive tests may be a factor that influences some test takers' perceptions of authenticity, and hence the way they respond to test items' (1990: 151). Characteristics of adaptive tests are the absence of feedback so test takers have no knowledge of the relevance or correctness of their responses to test items, and the influence of response on subsequent input. Oral interviews are often adaptive, and also reciprocal. Interviewees had different perceptions of the authenticity of tasks in the IELTS test according to their specific target situation. Only writing and reading were mentioned as being similar to academic study in the case of the IELTS test:

- R Which parts of the test do you think are like what you have to do at the university in real-life?
- S2 Writing, how to write an essay, that's similar to university. What else? I think ...
- R Just writing is most like the real situation?

- S3 I don't think no any part that can relate to my course. It's very different.
- R Would you like to explain in what way it's different?
- S3 For instance, writing, yes, sometimes it's a bit use some skill, for instance, introduction, body, and conclusion, but in my real course, my assignment, then you can say it's different from the IELTS test, and listening, listening is just the general situation, but in my class we talk know before.
- S1 I think reading and writing is like real situation.
- S4 I can compare to experience. In IELTS test we practice how to describe a process. That is exactly like our course, process, and that's like in IELTS test.

In contrast, the interviewees described the TOEFL test as 'general' with an emphasis on grammar and vocabulary, rather than including test tasks connected to target academic study. This perception appeared to be based on an assumption that the purposes of the TOEFL and the IELTS tests are different, as indicated above.

Final Comments

The final comments made by the focus group in the interview suggested that the IELTS test, which is a relatively new language proficiency test, is less internationally recognized than the TOEFL test. The opportunities for taking the test are less frequent and test centres are less widespread. Generally, the IELTS test was perceived to be a 'microcosm' of the target situation of academic study at the tertiary level in Australia. Subject-specific modules for the Reading and Writing subtests will inevitably favour some test takers with the relevant background knowledge. However, the integration of reading texts and writing tasks was perceived by the students as overcoming this problem to some extent. When comparing the TOEFL and the IELTS tests, the different weightings given to grammar as a test component affected students' perceptions of the nature of language proficiency testing and the relevance

of the test for future academic studies. An issue raised in the interview was the objectivity of testing in the TOEFL test compared with greater subjectivity in the IELTS test, particularly in the speaking subtest. Furthermore, there was a perception that significantly different test taking strategies were required for each language proficiency test.

Limitations of the Data Analysis and Interpretation

The research findings reported in this chapter were limited by the small size of the sample ($n=40$) and were biased towards the perceptions of Asian students. In addition, the focus group interview reflected only the viewpoints of five international students. The questionnaire survey revealed some problems of design. Questions were simplified, and may not have been sufficiently explicit, to take account of the fact that respondents were answering in a second language. Consequently, some responses to perceptions of test content tended to be ambiguous and data gathered about the IELTS test as a predictor of academic success may be unreliable. Interview responses indicated that the issue of test authenticity needed further clarification and exploration.

Finally, the results of a Chi-Square test cannot be reliably applied to a response data set unless certain assumptions are met (Brown, 1988; Woods *et al.*, 1986). All the *expected* frequencies in each contingency table must be higher than 5 for studies in which degrees of freedom (df) are less than 2. The contingency tables shown above fulfill this assumption (see Tables 5.5, 5.7, 5.9 and 5.10). However, due to the small sample size, there were limitations to the statistical analysis and to generalizability.

Summary

An investigation into international students' perceptions of the 1989 version of the IELTS test was undertaken at a significant period in its history before the introduction of a revised and updated 1995 version of the IELTS test. The research findings showed that international students perceived the 1989 version of the IELTS test to be relevant to performance in the target situation. It appeared to have 'test appeal' and to be acceptable to the test takers as an EAP/ESP test; in other words, it had high face validity. The IELTS test was compared with another internationally recognized and standardized language proficiency test, the TOEFL. Significant differences were identified between the two tests by international students. Furthermore, analysis and interpretation of the research findings raised questions about the appropriateness of an IELTS proficiency level of Band 6 as a recommended entry requirement for tertiary institutions. These issues, and the implications for the test takers of a revised version of the IELTS test, are discussed further in Chapter Six. Tentative conclusions are drawn and recommendations are made.

Chapter 6

Tentative Conclusions and Recommendations

Introduction

The possible implications for test takers of changes to the IELTS test from April 1995 are examined in the light of the findings reported in Chapter Five. Research findings based on a small data set do not offer conclusive evidence about international students' perceptions of the IELTS test. However, within the limitations of the small-scale survey undertaken, tentative generalizations are made about the face validity and the washback effects of the 1989 version of the IELTS test. The appropriacy of a threshold IELTS test proficiency level for entry to tertiary institutions in Australia is assessed. Recommendations are then made for the cautious interpretation of language proficiency test results by test users, and for the role of both intensive English language courses and English language support courses in the target academic setting. The final summary suggests areas of further research.

Research Findings and the IELTS Test Changes

A significant change to the IELTS test in 1995 is the reduction in ESP modules to one Academic module on the grounds that test centres and test users are often unclear about the appropriate test module for particular academic courses (UCLES, 1994a: 3). International students intending to undertake postgraduate studies, for example, are sometimes perplexed as to whether to choose a module based on their previous or future academic specialization. In addition, 75% of IELTS tests taken worldwide are module C, which may have influenced the decision to offer only one Academic module in a revised test (UCLES, 1994a).

Another reason for offering one Academic module is the claim that this does not favour groups of test takers in particular academic disciplines. Research into background knowledge as a potential source of test bias, discussed in Chapter Three, has tended to show that the performance of international students on reading comprehension tests is not only affected by language proficiency, but also by background knowledge in a familiar content area (Alderson and Urquhart, 1985a; 1985b). However, the task of designing reading comprehension tests that are academic but 'neutral' in content, without favouring students from a particular discipline, is challenging. As explained in Chapter Two, the Academic module also needs to be distinguishable in context from the General Training module which serves a particular test purpose in Australia.

The data analysed in Chapter Five suggested that the majority of international students perceived background knowledge in a specific subject to be helpful in both the Reading and Writing subtests of the 1989 version of the IELTS test. In addition, students' perceptions of the integration of the reading and writing components were generally favourable. Integration is an authentic simulation of the target situation; therefore, independent testing of the skills of reading and writing is a significant revision. In short, by effectively abandoning ESP testing, the test designers have redefined the construct of language proficiency underlying the revised IELTS test.

A further significant change in the IELTS test is the increase in test time by twenty minutes, and therefore the test is closer in length to the 1980 ELTS test described in Chapter Two. The research findings in this study showed that some international students perceived test time to be inadequate for the Reading and Writing subtests, and therefore extending test time may reduce the pressure of 'speededness'. However, it is also to be noted that the expected word count for each writing task has been increased to strengthen the Writing subtest.

Improving the reliability of the Writing and Speaking subtests by double marking scripts and checking the oral interview tapes of borderline candidates should increase perceptions of the fairness of the IELTS test. In this study, focus group interviewees drew attention to the objectivity of TOEFL test scores compared with the subjective scoring of parts of the IELTS test. The assessment of speaking and writing in the IELTS test is relevant to the target situation, but subjectivity in scoring raises the issue of inter-rater reliability. Consequently, any effort to improve accuracy in scoring, through appropriate training of examiners and checking of an individual candidate's diagnostic profile, will contribute to test reliability. Success or failure on the IELTS test can have major implications for the future of international students:

For example, a candidate profile of Reading 7; Listening 7; Writing 7; Speaking 4 would result in a final band score of 6. Given that 6.5 is the standard entry requirement for tertiary education, the subtest score of 4 for speaking would have impacted adversely on this candidate's chance of going to university (UCLES, 1994b: 3).

It should be noted that band score 6.5 on the IELTS test is the standard entry requirement for tertiary level studies in the UK and for many universities in Australia.

Tentative Generalizations

Due to the small-scale nature of the survey of international students' perceptions of the IELTS test undertaken as part of this thesis, it is difficult to make significant generalizations applicable to a wider target population. Perceptions of language proficiency tests may be affected by first language, cultural background, level of study, time spent in the target culture, language learning experiences, test wiseness, background knowledge, test

performance and other variables which are difficult to isolate even within large samples. Generalizations are necessarily tentative when research is based on a sample of 40 test takers.

Both the questionnaire and the focus group interview showed that international students perceived the IELTS test to be academically orientated, EAP in content, and a test of not only the four macroskills but also of grammar accuracy and lexis. Perceptions of the role of background knowledge in the ESP components of the test were inconclusive. It seems that unless an infinite number of ESP modules are offered in the IELTS test to encompass all academic disciplines, reading comprehension texts, and to a lesser extent writing tasks, will inevitably favour some test takers and not others. However, the integration of reading comprehension texts and writing tasks may compensate for some perceived unfairness in ESP testing.

The general opinion of international students in this study was that Band 6 on the IELTS test is only realistic as a baseline or minimum entry requirement for undergraduate level studies, and that a higher overall band score is certainly preferable for postgraduate studies. All four macroskills tested in the IELTS were perceived to be relevant to the target situation. However, their relative importance depended on the course of study undertaken. Minimum proficiency levels on the subtests of Reading and Writing for some linguistically demanding courses might be necessary. In other words, students prefer to be realistically informed of the demands of academic courses.

Whilst responses to the questionnaire survey and interview suggested that most international students believed that the IELTS test predicts readiness for academic study, they also perceived other factors to be influential. A sound academic background was one factor. A weak academic background makes the adjustment to academic study in Australian universities and colleges difficult as international students are expected to study independently and there is little place for rote learning, which might have been relevant in prior learning experiences.

Other factors identified as affecting academic success were hard work, motivation and verbal participation in class. Asian students have a reputation for conscientiousness, but for cultural and linguistic reasons they are generally reticent to ask questions in class and to express disagreement or to criticize. Social interaction outside the classroom was another factor identified as influencing academic success. Students also acknowledged the rapid improvement in their communicative ability as a result of immersion in the host culture. In addition, it was suggested that country of origin may affect ability to succeed in the target situation. International students who come from countries where English is used as an official language or for educational purposes tend to have greater confidence in the second language.

The research findings revealed that international students who had experience of taking both the TOEFL and the IELTS test were able to make perceptive comparisons. The TOEFL test was perceived to be more predictable, less orientated towards a specific target situation, and a more objective language proficiency test than the IELTS. However, a negative washback effect was recognized, namely acquiring ability in the test method. In addition, the TOEFL test was generally perceived to require of test takers a high level of grammar accuracy and lexis, whereas the IELTS test was perceived to require communicative language skills since it attempts to authentically simulate the target situation. The researcher feels that the Listening subtest is perhaps less successful in this regard.

Although based on a small sample of test takers, the research findings reported in this study point to the different reasons why the TOEFL and the IELTS tests had face validity. In the case of the TOEFL test, this seemed to be due to its perceived objectivity and to the fact that it is widely available and well-established amongst test users. In the case of the IELTS test, it had face validity since it was perceived to be a microcosm of the EAP target situation and because test preparation is also preparation for academic study. The IELTS and the TOEFL tests reflect different views of the construct of language proficiency, include different test

method facets and yet both have high face validity. The TOEFL test is essentially a general language proficiency test that is not intended to be an authentic simulation of an academic target situation. Test takers may choose not to have their oral ability assessed since the Test of Spoken English is a separate test. As the Test of Written English is optional, test takers' writing skills are not necessarily assessed through extended academic discourse. On the other hand, the TOEFL test does claim to be an internationally recognized, objective, reliable general proficiency test.

Research into the face validity of the IELTS test provided an insight into international students' perceptions of the test, how they prepare for the test, how they respond to test methods, and the extent to which they apply their language skills in the target situation. Research studies cited in Chapter Three showed that enhancing the face validity of a test may motivate test takers to perform well and to prepare for the target situation.

Tentative Conclusions

The study reported here has attempted an analysis of the evolution of an internationally recognized second language proficiency test, the IELTS. It is still a relatively new test, internationalized since 1989 in a joint project between Australia and the United Kingdom. At present, 35,000 candidates take the IELTS test annually, a small number compared with 780,000 candidates annually taking the TOEFL test. With the rapid increase in international students studying overseas, particularly in Australia, the IELTS test will gain more widespread recognition in future. However, it is clearly different in ideology and appearance from the TOEFL test which is more widely available and well-established.

Five years after its inception, the IELTS test has undergone a major revision and the nature of language proficiency testing underlying the test is no longer ESP/EAP but exclusively EAP.

The implications of changes to the IELTS test are significant not only for test takers but also for test users who judge the readiness of international students for academic study in the overseas target situation. Thus, the interpretation of IELTS test band scores requires cautious and informed decision-making by test users to ensure a closer match between minimum English language proficiency entry requirements and the linguistic demands of academic courses.

Investigating the face validity of a language proficiency test can verify the acceptability of a test internationally and its relevance to test takers and test users. Face validity needs consideration in test development. Therefore, consulting the test takers about their perceptions of an internationally recognized language proficiency test can contribute to a greater understanding of the factors affecting test performance, both linguistic and non-linguistic. Research into face validity may be a step towards improving the link between teaching, testing and target situation.

Recommendations to Test Users

There are an increasing number of language proficiency tests facing prospective international students. How do they choose the most appropriate proficiency test for the intended target situation? Is it determined by the availability of a test centre, or the advice of teachers, peers, or agents for overseas institutions? Moreover, does the preferred language proficiency test tend to reflect the existing EFL pedagogy in the non-English speaking environment? For example, communicative English language teaching methodology is probably still the exception rather than the norm in most parts of Asia. Consequently, tests favouring multiple-choice methods and receptive skills, with an emphasis on grammar, will closely match the learning style familiar to many prospective international students. Furthermore, test takers are

well aware that certain test methods, such as multiple-choice, can be mastered with practice, and that test wiseness increases the chances of success on proficiency tests.

As the demand for studying overseas grows, more and more prospective international students attend intensive English language courses in their home countries to prepare for language proficiency tests. Unless students are more effectively informed of the long term consequences of preparing for these tests, rather than simply the short term goals, they may experience problems and frustrations when studying in an English medium setting. Firstly, if international students are accepted into tertiary institutions with minimum language proficiency levels on tests that are not internationally recognized, their future academic options and level of study may be limited, or their opportunity to transfer between tertiary institutions may be restricted. Secondly, many students enter ELICOS courses with unrealistic expectations of the time needed to prepare for a language proficiency test, such as the IELTS test, or of the number of hours of English language instruction necessary to improve their overall band score result. Thirdly, the financial burden of misinformed choices can be considerable. These consequences frequently arise in the target situation causing dissatisfaction and disillusionment amongst international students. According to Weir (1993: 6), 'everything that students do to prepare themselves for the proficiency test should be contributory to enhanced performance in the future target situation'.

Many international students planning to study overseas have little experience of a teaching methodology that prepares them for the genre of western style academic study. Therefore, test users, namely admissions officers and academic departments within tertiary institutions, need to interpret the results of language proficiency tests cautiously. When interpreting the results of the IELTS test, for instance, it is essential that test users examine international students' diagnostic profiles of test results as well as overall band scores, and realistically advise them about the linguistic demands of academic courses. Minimum band score results

on specific macroskills are only a requirement for postgraduate studies and particular undergraduate courses in some universities.

As this study showed, international students may find the new academic genre daunting and the approach to research quite alien in the target situation. Test users should also be aware of the EFL pedagogy within international students' home countries and the problems of adjustment to a western style academic genre that international students will experience. As a result, the role of English language bridging courses in tertiary institutions is crucial. They offer the opportunity for students to acquire essential skills for the English-medium target situation, to have access to the available information technology, and to ease into the unfamiliar academic genre that rewards independent study, critical thinking and challenging the 'experts'.

The requirements of the target situation place new and challenging demands on international students. Even for those students who had the benefit of preparing for an EAP second language proficiency test, such as the IELTS, which is communicative in nature and which attempts to simulate the target situation, the advantages of English language support courses are evident. English language support courses have a complementary role, especially for students at the postgraduate level. Preparation for the IELTS test does not directly prepare students to participate in seminars, give oral presentations, develop academic notetaking skills and write research proposals. However, international students who undertake ELICOS courses in Australia before commencing academic studies are generally at a distinct advantage since many of these skills are developed in the language programmes offered.

Finally, research has convincingly shown that there are factors besides language proficiency level that may determine the ability of international students to cope in the target situation. This study showed that test takers perceive academic background, motivation, previous exposure to the second language, level of social interaction in the host country; and active

participation in classes to be significant factors. Those responsible for making decisions about the admission of international students into tertiary institutions in Australia need to obtain detailed profiles of students before determining their readiness for academic study. Language proficiency results seem to be only one component of these profiles. Gibson and Rusek (1992: 58) identified two further variables affecting academic success, educational and emotional support available at tertiary institutions, and thus they acknowledged the role of English language support courses and counselling services.

Language testing occurs in an educational and social setting, and the uses of language tests are determined largely by political needs that change over time and vary from one society to another. We must consider the value systems that inform test use-values of test developers, test takers, test users, the educational system, and society at large. We must consider the consequences, both positive and negative, of testing, as well as the relative value attached to these consequences (Bachman, 1990: 291).

Further Research

A valid methodology for investigating international students' perceptions of language proficiency tests needs to be established before this study can be broadened to include a wider international student population, representative of more nationalities, studying at a large number of tertiary institutions. Firstly, the response rate to mailed questionnaires is frequently disappointing. There are many factors that may deter respondents from completing and returning a questionnaire, such as English language level, information overload and number of questions, or the feeling that their responses will not be influential. Those who respond may hold stronger viewpoints and therefore threaten the validity of research findings. On the other hand, cultural factors may inhibit some international students from

openly expressing their attitudes towards a public examination, or their comments may be influenced by their test performance.

Secondly, it is sometimes suggested that attitudes and perceptions can only effectively be gathered using the L1 (first language) of respondents. However, the researcher agrees with Underhill (1987: 106) that even in the L2 (second language) test takers 'generally produce very informative and objective comments about tests, irrespective of their own personal performance'. The research findings in this study showed that international students can be thoughtful and forthcoming in their comments about language proficiency tests, possibly because they are rarely asked for their opinions about tests, and possibly because this study was undertaken independently of any examining body.

An effective methodology would be to distribute questionnaires or conduct interviews immediately after test takers have completed the IELTS test, either overseas or at ELICOS centres in Australia. This would probably improve the questionnaire response rate and improve the validity of data collected. Then students are not being asked to recall a test which they took some weeks, months or even years previously. The respondents would also include those test takers who fail to attain the minimum entry requirements for tertiary academic study in Australia and who are therefore judged to be academically unsuccessful, or diagnosed as in need of further English language instruction. In this study the sample of test takers (n=40) included international students who had taken an IELTS test and who had been accepted for academic courses. They were therefore academically successful if this is measured by meeting the second language proficiency level requirements of the university and by being admitted to an academic course. Therefore, the sample is more truly representative of the accessible population rather than the target population.

Nevertheless, the use of introspection techniques to establish test takers' strategies and attitudes towards test methods would necessarily exclude survey questions related to the

target situation. These questions provided a useful insight into test takers' perceptions of the IELTS test as a microcosm of the *real-life* academic environment, and indicated the appropriateness of test preparation for the target situation. The issues of relevance and appropriateness are pertinent to the research methodology since they affect the face validity of a test.

Introspection studies have shown that most language behaviour is only retained in short-term memory and that delayed retrospection is marginally less reliable than immediate retrospection (Cohen, 1987: 84). If this research is also applicable to test takers' perceptions of language proficiency tests, as well as to their responses to test items, then whether they took the test one week ago or one month ago may make little difference to their memory recall and to the reliability of the research findings. In other words, the effect of undertaking research that asks international students to recall a test taken some time previously may not invalidate the results. However, this is simply speculation and can best be validated by replicating the research at a later date with the same respondents.

Large samples of test takers representative of all international students studying in Australia, by nationality and first language, are needed to establish an appropriate methodology for assessing perceptions of the IELTS test. It might then be possible to effectively investigate different cultural perceptions of communicative language testing and test methods, supported by introspective studies that reveal how different cultural groups respond to various test methods.

A further area of research linked to this study is establishing the washback validity of the new version of the IELTS test. 'Changing a test is possibly the most powerful means of bringing about improvements in the learning experiences that go before it.' (Weir, 1993: 6). It is important to assess the impact of a new language proficiency test on English language learning internationally. The monitoring of test preparation in classrooms will establish the

potential for positive washback effects. One positive washback effect of the IELTS test might be the teaching of language skills and study skills necessary for international students to succeed in the target situation. Another positive effect of the test might be a more communicative language teaching methodology in the non-English speaking setting that balances the development of both receptive and active skills, and enables students to communicate more effectively in the target situation. As Weir (1993: 40) pointed out:

The most beneficial effect of a language proficiency test which assesses international students' readiness for English-medium academic study is that test preparation is also preparation for the target situation.

A second language proficiency test that closely resembles the target situation will have greater potential for positive washback. If the language test is highly predictable, however, the washback effects on teaching and learning will reflect this. Does preparation for different proficiency tests affect academic success in the target situation? Hughes supported Weir's viewpoint that preparation for a language proficiency test should also prepare students for the target situation. He stated that 'we cannot afford not to introduce a test with a powerful beneficial washback effect' (Hughes, 1989: 47). There is scope for research that assesses the washback effects on teaching and learning of different language proficiency tests and establishes whether future academic success is affected by these washback effects. In summary, areas of potential research are:

- further investigations into test takers' perceptions of proficiency tests,
- investigations into test users' perceptions of proficiency tests,
- investigations into the washback effects of proficiency tests, and
- introspection techniques to determine the strategies that test takers' use.

Bibliography

- Adams, R.J., P.E. Griffin and L. Martin. 1987. A latent trait method for measuring a dimension in second language proficiency. *Language Testing*, 4, 1, pp. 9-27.
- Alderson, J.C. 1981. Reaction to the Morrow Paper. In J.C. Alderson and A. Hughes. 1981, pp. 45-54.
- Alderson, J.C. 1986. Innovations in Language Testing? In M. Portal. 1986, pp. 93-105.
- Alderson, J.C. 1988a. New procedures for validating proficiency tests of ESP? Theory and practice. *Language Testing*, 5, 2, pp. 220-232.
- Alderson, J.C. 1988b. Testing English for Specific Purposes: How Specific Can We Get? In A. Hughes. 1988a, pp. 16-28.
- Alderson, J.C. 1991a. Language Testing in the 1990s: How Far Have We Come? How Much Further Have We To Go? In S. Anivan. 1991a, pp. 1-26.
- Alderson, J.C. 1991b. Bands and Scores. In J.C. Alderson and B. North. 1991, pp. 71-86.
- Alderson, J.C. and C. Clapham. 1992. Applied linguistics and language testing: A case study of the ELTS test. *Applied Linguistics*, 13, 2, pp. 149-167.
- Alderson, J.C. and A. Hughes. (eds). 1981. *Issues in Language Testing*. ELT Documents 111. London: The British Council.
- Alderson, J.C. and B. North. 1991. *Language Testing in the 1990s: The Communicative Legacy*. London: Macmillan.
- Alderson, J.C. and A.H. Urquhart. 1985a. This Test is Unfair: I'm not an Economist. In P.C. Hauptman, R. LeBlanc and M.B. Wesche. 1985, pp. 25-43.
- Alderson, J.C. and A.H. Urquhart. 1985b. The effect of students' academic discipline on their performance on ESP reading tests. *Language Testing*, 2, 2, pp. 192-204.

- Alderson, J.C. and D. Wall. 1993. Does washback exist? *Applied Linguistics*, 14, 2, pp. 115-129.
- Allan, A. 1992. Development and validation of a scale to measure test-wiseness in EFL/ESL reading test takers. *Language Testing*, 9, 2, pp. 101-122.
- Anastasi, A. 1982. *Psychological Testing*. Fifth Edition. London: Macmillan.
- Anivan, S. (ed.). 1991a. *Current Developments in Language Testing*. Anthology Series 25. Singapore: SEAMEO RELC.
- Anivan, S. (ed.). 1991b. *Issues in Language Programme Evaluation in the 1990s*. Anthology Series 27. Singapore: SEAMEO RELC.
- Ary, D., L.C. Jacobs and A. Razavieh. 1979. *Introduction to Research in Education*. Second Edition. New York: Holt, Rinehart and Winston.
- Bachman, L.F. 1988. Language testing-SLA research interfaces. *Annual Review of Applied Linguistics*, Vol. 9, pp. 193-209.
- Bachman, L.F. 1989. The Development and Use of Criterion-referenced Tests of Language Ability in Language Program Evaluation. In R.K. Johnson. 1989, pp. 242-258.
- Bachman, L.F. 1990. *Fundamental Considerations in Language Testing*. Oxford: OUP.
- Bachman, L.F. 1991. What does language testing have to offer? *TESOL Quarterly*, 25, 4, pp. 671-704.
- Bachman, L.F. and John L.D. Clark. 1987. The measurement of foreign/second language proficiency. *Annals of the American Academy of Political and Social Science*, 490, pp. 20-33.
- Bachman, L.F., A. Kunnan, S. Vanniarajan and B. Lynch. 1988. Task and ability analysis as a basis for examining content and construct comparability in two EFL proficiency test batteries. *Language Testing*, 5, 2, pp. 128-159.

- Bachman, L.F. and A.S. Palmer. 1986. The construct validation of self-ratings of communicative language ability. *Language Testing*, 6,1, pp. 14-29.
- Bachman, L.F. and S.J. Savignon. 1986. The evaluation of communicative language proficiency: a critique of the ACTFL Oral Interview. *The Modern Language Journal*, 70, 4, pp. 380-390.
- Baker, D. 1989. *Language Testing*. London: Edward Arnold.
- Ballard, B. 1987. Academic adjustment: the other side of the export dollar. *Higher Education Research and Development*, 6, 2, pp. 109-119.
- Ballard, B. and J. Clanchy. 1984. *Study Abroad. A Manual for Asian Students*. Kuala Lumpur: Longman.
- Ballard, B. and J. Clanchy. 1988. *Studying in Australia*. Melbourne: Longman Cheshire.
- Ballard, B. and J. Clanchy. 1991. *Teaching Students from Overseas*. Melbourne: Longman Cheshire.
- Berk, R.A. (ed.). 1984. *A Guide to Criterion-Referenced Test Construction*. Baltimore, Md.: The John Hopkins University Press.
- Bers, T.H. 1987. Exploring institutional images through focus group interviews. *New Directions for Institutional Research*, No. 54, 14, 2, pp. 19-29.
- Brindley, G. 1986. *The Assessment of Second Language Proficiency: Issues and Approaches*. NSCR Research Series. Adelaide: National Curriculum Resources Centre.
- Brindley, G. 1989. *Assessing Achievement in the Learner-Centred Curriculum*. Sydney: NCELTR, Macquarie University.
- Brodigan, D.L. 1992. *Focus Group Interviews: Applications for Institutional Research*. Association for Institutional Research Professional File, No. 43.

- Brown, J.D. 1981. Newly Placed Students versus Continuing Students: Comparing Proficiency. In J.C. Fisher, M.A. Clarke and J. Schachter. (eds). 1981. *Research and Practice in Teaching English as a Second Language*. Washington, DC.: TESOL, pp. 111-119.
- Brown, J.D. 1988. *Understanding Research in Second Language Learning*. Cambridge: CUP.
- Brown, J.D. 1990. Short-cut estimators of criterion-referenced test consistency. *Language Testing*, 7,1, pp. 77-97.
- Brown, J.D. 1991. Statistics as a foreign language- part 1: what to look for in reading statistical language studies. *TESOL Quarterly*, 25, 4, pp. 569-585.
- Brown, J.D. 1992. Statistics as a foreign language- part 2: more things to consider in reading statistical language studies. *TESOL Quarterly*, 26, 4, pp. 629-664.
- Brown, J.D. (forthcoming). *Testing in Language Programs*.
- Buck, G. 1992. Listening comprehension: construct validity and trait characteristics. *Language Learning*, 42, 3, pp. 313-357.
- Burns, R. 1991. Study and stress among first year overseas students in an Australian university. *Higher Education Research and Development*, 10, 1, pp. 61-77.
- Busch, M. 1993. Using Likert scales in L2 research. A researcher comments. *TESOL Quarterly*, 27, 4, pp. 733-736.
- Canale, M. 1988. The measurement of communicative competence. *Annual Review of Applied Linguistics*, vol. 8, pp. 67-84.
- Canale, M. and M. Swain. 1980. Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1, 1, pp. 1-47.
- Carey, M.A. 1994. The Group Effect in Focus Groups: Planning, Implementing, and Interpreting Focus Group Research. In J.M. Morse. (ed.). 1994. *Critical Issues in Qualitative Research Methods*. California: Sage Publications, pp. 225-242.

- Carroll, B. 1983. The English Language Testing Service. In R.R. Jordan. (ed.). 1983. *Case Studies in ELT*. London & Glasgow: Collins ELT, pp. 90-101.
- Chapelle, C. 1988. Field independence: A source of language test variance? *Language Testing*, 5, 1, pp. 62-82.
- Chen, Z. and G. Henning. 1985. Linguistic and cultural bias in language proficiency tests. *Language Testing*, 2, 2, pp. 155-163.
- Cohen, A. 1984. On taking language tests: what the students report. *Language Testing*, 1, 1, pp. 70-81.
- Cohen, A. 1987. Using Verbal Reports in Research on Language Learning. In C. Faerch and G. Kasper. 1987, pp. 82-95.
- Cohen, L. and L. Manion. 1980. *Research Methods in Education*. London: Croom Helm.
- Cooper, R.L. 1986. Panel discussion: the next 25 years? *Language Testing*, 3, 2, pp. 239-240.
- Criper, C. and A. Davies. 1988. *ELTS Validation Project Report*. ELTS Research Report 1(i). London: The British Council and the University of Cambridge Local Examinations Syndicate.
- Davies, A. 1983. The Validity of Concurrent Validation. In A. Hughes and D. Porter. 1983, pp. 141-147.
- Davies, A. 1984. Validating three tests of English language proficiency. *Language Testing*, 1, 1, pp. 50-69.
- Davies, A. 1986. Indirect ESP Testing: Old Innovations. In M. Portal. 1986, pp. 55-67.
- Davies, A. 1988a. Operationalizing uncertainty in language testing: An argument in favour of content validity. *Language Testing*, 5, 1, pp. 32-48.
- Davies, A. 1988b. Communicative Language Testing. In A. Hughes. 1988a, pp. 5-15.

- Davies, A. 1988c. Procedures in Language Test Validation. In A. Hughes. 1988a, pp. 29-35.
- Deakin, G. 1994. *Practice Tests for IELTS*. Victoria: Indonesia Australia Language Foundation and Hawthorn Institute of Education.
- DEET. 1993. *Overseas Student Statistics*. Canberra: Australian Government Publishing Service.
- Elder, C. 1993. Language proficiency as a predictor of performance in teacher education. *Melbourne Papers in Language Testing*, 2, 1, pp. 68-89.
- Educational Testing Service. 1994. *Bulletin of Information for TOEFL, TWE, and TSE*. Princeton, NJ: Educational Testing Service.
- Faerch, C. and G. Kasper. (eds). 1987. *Introspection in Second Language Research*. Clevedon, UK: Multilingual Matters.
- Foddy, W. 1993. *Constructing Questions for Interviews and Questionnaires*. Cambridge: CUP.
- Fowler, F.J. Jr. and T.W. Mangione. 1990. *Standardized Survey Interviewing Minimizing Interviewer-Related Error*. Newbury Park, California: Sage Publications.
- Gay, L.R. 1992. *Educational Research. Competencies for Analysis and Application*. Fourth Edition. New York: Merrill.
- Gibson, C. and W. Rusek. 1992. *The validity of an overall band score of 6.0 on the IELTS test as a predictor of adequate English language level appropriate for successful academic study*. Unpublished MA dissertation, Macquarie University.
- Graham, J.G. 1987. English language proficiency and the prediction of academic success. *TESOL Quarterly*, 21, 3, pp. 505-521.
- Griffin, P. 1990. Characteristics of the test components of the IELTS battery: Australian trial data. Paper presented at *RELC Annual Seminar*, Singapore.

- Grotjahn, R. 1987. On the Methodological Basis of Introspection Methods. In C. Faerch and G. Kasper. 1987, pp. 54-81.
- Hamilton, J., M. Lopes, T. McNamara and E. Sheridan. 1993. Rating scales and native speaker performance on a communicatively oriented EAP test. *Melbourne Papers in Language Testing*, 2, 1, pp. 1-24.
- Hamp-Lyons, L. 1989. Applying the partial credit method of Rasch analysis: language testing and accountability. *Language Testing*, 6, 1, pp. 109-118.
- Hatch, E. and H. Farhady. 1982. *Research Design and Statistics for Applied Linguistics*. Rowley, Mass.: Newbury House.
- Hauptman, P.C., R. LeBlanc and M.B. Wesche. (eds). 1985. *Second Language Performance Testing*. Ottawa: University of Ottawa Press.
- Heaton, J.B. 1988. *Writing English Language Tests*. London: Longman.
- Henning, G. 1984. Advantages of latent trait measurement in language testing. *Language Testing*, 1, 2, pp. 123-133.
- Henning, G. 1987. *A Guide to Language Testing*. Cambridge, Mass.: Newbury House.
- Henning, G. 1989. Comments on the comparability of TOEFL and Cambridge CPE. *Language Testing*, 6, 2, pp. 217-222.
- Henning, G., T. Hudson and J. Turner. 1985. Item response theory and the assumption of unidimensionality for language tests. *Language Testing*, 2, 2, pp. 141-154.
- Hudson, T. 1991. Relationships among IRT item discrimination and item fit indices in criterion-referenced language testing. *Language Testing*, 8, 2, pp. 160-181.
- Hudson, T. 1989. Mastery Decisions in Program Evaluation. In R.K. Johnson. 1989, pp. 259-269.
- Hudson, T. and B. Lynch. 1984. A criterion-referenced measurement approach to ESL achievement testing. *Language Testing*, 1, 2, pp. 171-201.

- Hughes, A. 1986. A Pragmatic Approach to Criterion-referenced Foreign Language Testing. In M. Portal. 1986, pp. 31-40.
- Hughes, A. (ed.). 1988a. *Testing English for University Study*. ELT Documents 127. Oxford: Modern English Publications and The British Council.
- Hughes, A. 1988b. Introducing a Needs Based Test of English for Study in an English Medium University in Turkey. In A. Hughes. 1988a, pp. 134-153.
- Hughes, A. 1989. *Testing for Language Teachers*. Cambridge: CUP.
- Hughes, A. and D. Porter. (eds). 1983. *Current Developments in Language Testing*. London: Academic Press.
- Hyltenstam, K. and M. Pienemann. (eds). 1985. *Modelling and Assessing Second Language Acquisition*. Clevedon, UK: Multilingual Matters.
- IELTS. 1992. *An Introduction to IELTS*. Cambridge: The British Council, UCLES and IDP.
- IELTS Australia. 1994. Report IELTS Administrators Meeting. University of Queensland.
- Ingram, D.E. 1984. *Australian Second Language Proficiency Ratings*. Canberra: Department of Immigration and Ethnic Affairs.
- Ingram, D.E. 1985. Assessing Proficiency: An Overview on Some Aspects of Testing. In K. Hyltenstam and M. Pienemann. 1985, pp. 215-276.
- Ingram, D.E. 1990. The International English Language Testing System (IELTS): The Speaking Test. Workshop presentation at *RELC Annual Seminar*, Singapore.
- Ingram, D.E. 1991. The International English Language Testing System (IELTS): Its Nature and Development. In S. Anivan. 1991a, pp. 185-199.
- Johns, A.M. and T. Dudley-Evans. 1991. English for specific purposes: international in scope, specific in purpose. *TESOL Quarterly*, 25, 2, pp. 297-314.

- Johnson, R.K. (ed.). 1989. *The Second Language Curriculum*. Cambridge: CUP.
- Lee, Y.P., Angela C.Y.Y. Fok, R. Lord and G. Low. (eds). 1985. *New Directions in Language Testing*. Hong Kong: Pergamon.
- Light, R.L., M. Xu and J. Mossop. 1987. English proficiency and academic performance of international students. *TESOL Quarterly*, 21, 2, pp. 251-261.
- Lynch, B., F. Davidson and G. Henning. 1988. Person dimensionality in language test validation. *Language Testing*, 5, 2, pp. 206-219.
- McNamara, T.F. 1990. Item Response Theory and the validation of an ESP test for health professionals. *Language Testing*, 7, 1, pp. 52-75.
- McNamara, T.F. 1991. Test dimensionality: IRT analysis of an ESP listening test. *Language Testing*, 8, 2, pp. 139-159.
- Milanovic, M. 1991. Materials-Based Tests: How Well Do They Work? In S. Anivan. 1991a, pp. 119-138.
- Mishler, E.G. 1986. *Research Interviewing*. Cambridge, Mass.: Harvard University Press.
- Morrow, K. 1986. The Evaluation of Tests of Communicative Performance. In M. Portal. 1986, pp. 1-13.
- Munby, J. 1978. *Communicative Syllabus Design*. Cambridge: Cambridge University Press.
- Nevo, B. 1985. Face validity revisited. *Journal of Educational Measurement*, 22, 4, pp. 287-293.
- Nevo, D. 1986. Panel discussion: the next 25 years? *Language Testing*, 3, 2, pp. 240-242.
- Nixon, U. 1993. Coping in Australia: problems faced by overseas students. *Prospect*, 8, 3, pp. 42-51.
- Nunan, D. 1988. *The Learner-Centred Curriculum*. Cambridge: CUP.

- Nunan, D. 1991. Second Language Proficiency Assessment and Program Evaluation. In S. Anivan. 1991b, pp. 46-59.
- Nunan, D. 1992. *Research Methods in Language Learning*. Cambridge: CUP.
- Oller, J. 1983. Response to Vollmer. 'g', what is it? In A. Hughes and D. Porter. 1983, pp. 35-37.
- Palmer, A.S. and L.F. Bachman. 1981. Basic Concerns in Test Validation. In J.C. Alderson and A. Hughes. 1981, pp. 135-151.
- Popham, W.J. 1978. *Criterion-Referenced Measurement*. Englewood Cliffs, NJ: Prentice-Hall.
- Portal, M. (ed.). 1986. *Innovations in Language Testing*. Windsor, Berks.: NFER-Nelson.
- Prapphal, K. 1990. *The Relevance of Language Testing Research in the Planning of Language Programmes*. Bangkok: Chulalongkorn University Language Institute, pp. 56-67.
- Read, J. 1991. The Validity of Writing Test Tasks. In S. Anivan. (ed.). 1991a.
- Reed, D.J. 1992. The relationship between criterion-based levels of oral proficiency and norm-referenced scores of general proficiency in English as a second language. *System*, 20, 3, pp. 329-345.
- Reid, J. 1990. The dirty laundry of ESL survey research. *TESOL Quarterly*, 24, 2, pp. 323-338.
- Reves, T. and A. Levine. 1992. From needs analysis to criterion-referenced testing. *System*, 20, 2, pp. 201-210.
- Richards, J., J. Platt and H. Weber. 1985. *Longman Dictionary of Applied Linguistics*. Harlow, England: Longman.
- Samuelowicz, K. 1987. Learning problems of overseas students: two sides of a story. *Higher Education Research and Development*, 6, 2, pp. 121-133.

- Savignon, S.J. 1983. *Communicative Competence: Theory and Classroom Practice*. Reading, Mass.: Addison-Wesley.
- Seaton, I. 1983. The English Language Testing Service (ELTS): Two Issues in the Design of the New 'Non-academic Module'. In A. Hughes and D. Porter. 1983, pp. 129-139.
- Shectman, Z. 1988. Selecting candidates for teachers' training college: a group assessment procedure as an alternative to individual interviews. *European Journal of Teacher Education*, 11, 2/3, pp. 185-193.
- Shohamy, E. 1984. Does the testing method make a difference? The case of reading comprehension. *Language Testing*, 1, 2, pp. 147-170.
- Shohamy, E. 1988. A proposed framework for testing the oral language of second/foreign language learners. *Studies in Second Language Acquisition*, 10, 2, pp. 165-179.
- Shohamy, E. and O. Inbar. 1991. Validation of listening comprehension tests: The effect of text and question type. *Language Testing*, 8, 1. pp. 23-40.
- Shohamy, E. and T. Reves. 1985. Authentic language tests: where from and where to? *Language Testing*, 2, 1, pp. 48-59.
- Skehan, P. 1984. Issues in the testing of English for specific purposes. *Language Testing*, 1, 2, pp. 202-220.
- Skehan, P. 1988. Language testing- part I. *Language Teaching Abstracts*, 21, 4, pp. 211-221.
- Skehan, P. 1989. Language testing- part II. *Language Teaching Abstracts*, 22, 1, pp. 1-13.
- Spolsky, B. 1985a. The limits of authenticity in language testing. *Language Testing*, 2, 1, pp. 31-40.
- Spolsky, B. 1985b. What does it mean to know how to use a language? an essay on the theoretical basis of language testing. *Language Testing*, 2, 2, pp. 180-191.

- Spolsky, B. 1986. A multiple choice for language testers. *Language Testing*, 3, 2, pp. 147-158.
- Spolsky, B. 1989. Communicative competence, language proficiency, and beyond. *Applied Linguistics*, 10, 2, pp. 138-156.
- Spolsky, B. 1990. The prehistory of TOEFL. *Language Testing*, 7, 1, pp. 98-118.
- Stevenson, D.K. 1985a. Authenticity, validity and a tea party. *Language Testing*, 2, 1, pp. 41-47.
- Stevenson, D.K. 1985b. Pop Validity and Performance Testing. In Y.P. Lee et al. 1985, pp. 111-119.
- Swain, M. 1985. Large-scale Communicative Language Testing: a case study. In Y.P. Lee et al. 1985, pp. 35-46.
- Swales, J.M. 1990. *Genre Analysis*. Cambridge: CUP.
- Taylor, D.S. 1988. The meaning and use of the term 'competence' in linguistics and applied linguistics. *Applied Linguistics*, 9, 2, pp. 148-168.
- Traynor, R. 1985. The TOEFL: an appraisal. *ELT Journal*, 39, 1, pp. 43-47.
- Turner, J. 1993. Another researcher comments. *TESOL Quarterly*, 27, 4, pp. 736-739.
- UCLES. 1994a. *IELTS 1995*. Information Circular. Cambridge: The British Council, UCLES and IDP.
- UCLES. 1994b. *Proposal for the Maintenance of Measurement Standards in IELTS*. Cambridge: The British Council, UCLES and IDP.
- Underhill, N. 1987. *Testing Spoken Language*. Cambridge: CUP.
- University of Adelaide. 1992. *Survey of Overseas Students' English Language Needs while Studying at Adelaide University*. Adelaide: Centre for Language Teaching and Research.

- Weir, C.J. 1983. The Associated Examining Board's Test of English for Academic Purposes: an Exercise in Content Validation. In A. Hughes and D. Porter. 1983, pp. 147-153.
- Weir, C.J. 1988. The Specification, Realization and Validation of an English Language Proficiency Test. In A. Hughes. 1988a, pp. 45-110.
- Weir, C.J. 1990. *Communicative Language Testing*. Hemel Hempstead: Prentice-Hall.
- Weir, C.J. 1993. *Understanding and Developing Language Tests*. Hemel Hempstead: Prentice-Hall.
- Westaway, G. 1988. Developments in the English Language Testing Service (ELTS) M2 writing test. *Australian Review of Applied Linguistics*, 11, 2, pp. 13-29.
- Widdowson, H.G. 1989. Knowledge of language and ability for use. *Applied Linguistics*, 10, 2, pp. 128-137.
- Woods, A., P. Fletcher and A. Hughes. 1986. *Statistics in Language Studies*. Cambridge: CUP.

Appendix I

THE INTERNATIONAL ENGLISH LANGUAGE TESTING SYSTEM

The IELTS test is divided into four subtests. Each candidate takes two general subtests, Listening and Speaking, and a module with two specialised subtests, Reading and Writing. The general subtests test general English while the modules test skills in particular areas suited to the candidate's chosen course of study.

The tests are taken in this order:

Reading	55 minutes
Writing	45 minutes
Listening	30 minutes
Speaking	11–15 minutes
Total Test Length –	2 hours 25 minutes

Reading and Writing

Four modules are available. These modules test reading and writing in an academic or general training context. Texts and graphic material are drawn from the candidates' broad field of study.

Module A – Physical Sciences and related disciplines

Module B – Biological Sciences and related disciplines

Module C – Humanities and Social Sciences and related disciplines

General Training Module – General or industrial training, short placement courses, upper secondary school programmes.

In the Reading section (55 minutes) the skills and functions tested include following instructions, finding main ideas, identifying the underlying concept or theme and identifying relationships between ideas in the text. Question-types include gap-filling, short-answer questions, sorting events into order, and multiple choice.

In the Writing section (45 minutes) the skills and functions tested include organising and presenting data, listing the stages of a procedure, describing an object or event or sequence of events, evaluating and challenging ideas, evidence and argument.

Candidates are asked to complete two tasks, one involving information transfer or reprocessing (15 minutes) and the other requiring candidates to draw on information from a variety of sources and their own experience to present an argument or discussion on a given topic (30 minutes).

Writing test papers are marked by qualified examiners, trained in IELTS test procedures. Ratings are monitored by UCLES on a regular basis to ensure that uniform standards are maintained.

Listening

This is a test of listening comprehension in the context of general English proficiency. The earlier sections deal with social survival topics – travel, accommodation, health and welfare – and the later ones with topics in the area of education and training. The test is based on a tape recording and item-types include controlled note-taking, form-filling, gap-filling and short-answer questions. As in the Reading subtest, extensive use of multiple choice questions is avoided.

Speaking

The Speaking Test is a structured oral interview, designed to encourage candidates to demonstrate their ability to speak English. The emphasis is on general speaking skills, rather than those related to students' own fields of study. The test is rated according to a global proficiency scale. In making their assessments, interviewers take account of effective communication and appropriate and flexible use of grammar and vocabulary. All interviews are recorded so that ratings given by interviewers can be monitored on a regular basis.

(IELTS, 1992: 4)

Appendix II

THE INTERNATIONAL ENGLISH LANGUAGE TESTING SYSTEM PROFICIENCY RATING SCALE

Overall Band Scores

The nine Bands and their descriptive statements are as follows:

Band 9 Expert User. Has fully operational command of the language: appropriate, accurate and fluent with complete understanding.

Band 8 Very Good User. Has fully operational command of the language with only occasional unsystematic inaccuracies and inappropriacies. Misunderstandings may occur in unfamiliar situations. Handles complex detailed argumentation well.

Band 7 Good User. Has operational command of the language, though with occasional inaccuracies, inappropriacies and misunderstandings in some situations. Generally handles complex language well and understands detailed reasoning.

Band 6 Competent User. Has generally effective command of the language despite some inaccuracies, inappropriacies and misunderstandings. Can use and understand fairly complex language, particularly in familiar situations.

Band 5 Modest User. Has partial command of the language, coping with overall meaning in most situations, though is likely to make many mistakes. Should be able to handle basic communication in own field.

Band 4 Limited User. Basic competence is limited to familiar situations. Has frequent problems in understanding and expression. Is not able to use complex language.

Band 3 Extremely Limited User. Conveys and understands only general meaning in very familiar situations. Frequent breakdowns in communication occur.

Band 2 Intermittent User. No real communication is possible except for the most basic information using isolated words or short formulae in familiar situations and to meet immediate needs. Has great difficulty understanding spoken and written English.

Band 1 Non User. Essentially has no ability to use the language beyond possibly a few isolated words.

Band 0 Did not attempt the test. No assessable information.

(IELTS, 1992: 6)

Appendix III

THE INTERNATIONAL ENGLISH LANGUAGE TESTING SYSTEM

INDEX OF ACADEMIC MODULES

Index of Graduate and Undergraduate Courses and appropriate IELTS Module to be taken:

Accountancy	C	Applied Genetics	B
Actuarial Science	C	Applied Geochemistry	A
Administration	C	Applied Linguistics	C
Administration in Nursing	B	Applied Mathematics	A
Agriculture	B	Applied Zoology	B
Agricultural Economics	B	Aquaculture	B
Agricultural Extension	B	Archaeology	C
Agri-Economics	C	Architecture	C
Agronomy	B	Architectural Engineering	A
Anaesthesiology	B	Architectural Science	A
Animal Nutrition	B	Art History	C
Animal Production	B	Audiology	B
Anthropology	C	Automotive Engineering	A
Applied Entomology	B	Aviation	A
<hr/>			
Bacteriology	B	Building Engineering	A
Biochemistry	B	Business Administration	C
Biology	B	Business Computing Systems	C
Biomechanics	B	Business Management	C
Biophysics	B	Business/Shipping	C
Biotechnology	B	Business Studies	C
<hr/>			
Cardiology	B	Computer Engineering	A
Cartographic Engineering	A	Computer Information Systems	C
Chemical Catalysts	A	Computer Science	A
Chemical Engineering	A	Computer Studies	C
Chemical Oceanography	A	Conservation of Plant	
Chemistry	A	Genetic Resources	B
Child Health	B	Conservation Science	B
Cinema/Video Preservation	C	Conservation of Soil Fertility	B
Civil Engineering	A	Construction Management	A
Clinical Medicine	B	Control Systems	A
Clinical Tropical Medicine	B	Co-operative Accounting	C
Cognitive Science	C	Co-operative Development and	
Commerce	C	Financial Management	C
Communications	C	Co-operative Education and	
Community Health	B	Training	C
Computer-Aided		Crop Physiology	B
Engineering	A		

Demography	C	Digital Communication	A
Dentistry	B	Diplomacy	C
Dermatology	B	Distribution Engineering	A
Design and Technology	A	Drama	C
Development Studies	C		
<hr/>			
Earth Science (Life)	C	Energy	A
Earth Science (Physical)	A	Energy Economics	C
Earth Resource Technology	A	Energy Engineering	A
Ecology	B	Engineering	A
Economics	C	Engineering Surveying	A
Education	C	English Language Teaching	C
Education by Radio and TV	C	English Literature	C
Education Planning	C	Entomology	B
Education Supervision/ Inspection	C	Environmental Engineering	A
Educational Research	C	Environmental Forestry	B
Educational Technology	C	Environmental Health	B
Electrical Engineering	A	Environmental Science	A
Electrical Engineering Optics	A	Enzymes	B
Electrical Measurement	A	Epidemiology	B
Electrical Power Engineering	A	Estate Management	C
Electronic Data Processing	A	Ethanol from Industrial and Agricultural Waste	A
Electro-Mechanical Engineering	A	European and EEC Law	C
Electronic Telecommunications	A	Experimental Psychology	C
		Exploration Geophysics	A
<hr/>			
Fermentation Technology	A	Fisheries Science	B
Film and Video	C	Fluid Mechanics	A
Finance	C	Food and Agricultural Biotechnology	B
Fine Arts	C	Food Engineering	B
Fisheries Planning and Management	B		
<hr/>			
Geology	A	Geography	C
Geophysics	A	Graphic Design	C
<hr/>			
History	C	Human Genetics	B
Horticulture	B	Human Nutrition	B
Hospital Architecture	A	Human Resources Development	C
Hotel and Catering Management	C	Hydrogeology	A
		Hydrology	A
Immunology	B	Instrumentation Digital Signal Processing	A
Industrial Design	C	Intensive Care Units	B
Industrial Project Planning	C	Intensive Therapy Paediatrics	B
Industrial Relations	C	International Economics	C
Industrial Safety	A	International Finance and Business	C
Industrialisation and Management	C	International Law	C
Infrastructure Planning	A	International Relations and Politics	C
Irrigation Engineering	A	International Studies	C
Instrumentation and Automatic Systems	A		
Intensive Care Units	B		

Journalism	C		
Labour Economics	C	Laser Spectroscopy	A
Land Economy	C	Law	C
Land Management	C	Library and Information Studies	C
Land Protection	B	Life Science	B
Land Surveying	C	Linguistics	C
Land Valuation	C	Loan Administration	C
Landscape Engineering	B	Local Government	C
Languages	C		
Laser Machining	A		
Machine Tool Manufacturing Technology	A	Meteorology	A
Management Studies	C	Metrication and Legal Metrology	C
Manpower Studies	C	Microbiology	B
Marine and Fisheries Science	B	Microelectronics	A
Maritime Studies	A	Mineral Economics	A
Marketing	C	Mineral Exploration	A
Materials Engineering	A	Mining Engineering	A
Materials Science	A	Museum Studies	C
Mathematics	A	Music	C
Mechanical Engineering	A	Music and Sound Recording	A
Media Studies	C	Music Technology	C
Medicine	B	Mycology	B
Metallurgy and Materials Science	A	Mycotoxins	B
National and Project Planning	C	Nuclear Physics	A
Naval Engineering	A	Nursing (General)	B
Neurology	B	Nursing (Specialised)	B
Neuro-Surgery	B	Nutrient Enrichment	B
Nuclear Medicine	B	Nutrition	B
Operational Research	C	Optometry	B
Ophthalmology	B		
Paediatric Medicine	B	Planning Studies	C
Palaetology	B	Plant Biology	B
Parasitology	B	Plant Biotechnology	B
Pathology	B	Plant Physiology	B
Periodontology	B	Plant Sciences	B
Personnel Management	C	Police Management	C
Pest Management	B	Political Economy	C
Petroleum Engineering	A	Political Science	C
Petroleum Geology	A	Politics	C
Pharmacy	B	Population Studies	C
Pharmaceutical Technology	B	Post-Harvest Technology	B
Pharmacology	B	Power Electronics	A
Philosophy	C	Power Engineering	A
Photographic Printing and Cameras	A	Power Station Maintenance	A
Photoionization of Molecules	A	Printing	A
		Production Engineering	A

Physical Chemistry	A	Production Management	C
Physical Education	C	Psychology	C
Physical Science	A	Public Administration	C
Physical Therapy	B	Public Order	C
Physics	A	Pure Mathematics	A
Physiology	B		
Quantity Surveying	C	Quarantine	B
Radiation	B	Religious Studies	C
Radiography	B	Rural Management	C
Railway Engineering	A	Rural Research and Policy	C
Range Ecology	B	Rural Social Development	C
Science Policy	C	Solar Energy	A
Science and Technology	A	Solar Radiation	A
Scientific Research	A	Solid State Physics	A
Seed Production	B	Solid Waste Disposal and Treatment	A
Ship Administration	C	Spectroscopy	A
Ship Construction and Repair	A	Speech Studies	B
Ship Science	A	Statistics	C
Shipping Law	C	Structural Vibrations	C
Social Anthropology	C	Supply and Material Management	C
Social and Political Science	C	Surgical Intensive Care	B
Sociology	C	Surveying	C
Soil Mechanics	A	System Protection	A
Soil Nutrition	B		
Solar Architecture	C		
Technology	A	Traffic Planning	C
Telecommunications	C	Translation and Linguistics	C
Textile Technology	A	Transport Design	C
Theatre Studies	C	Transport Economics	C
Theology	C	Transport Engineering	A
Timber Technology	A	Transport Management	C
Tissue Culture	B	Transportation	A
Tourism	C	Tropical Agriculture	B
Town Planning	C	Tropical Medicine	B
Toxicology	B	Tropical Veterinary Science	B
Traffic Engineering	C		
Urban Design	C	Urban Management	C
Urban Development	C	Urban Planning	C
Urban Land Appraisal	C	Urbanism	C
Veterinary Science	B	Virology	B
Water and Waste Engineering	A	Water Resources	A
Youth Leadership	C		
Zoology	B		

Appendix IV

INTERNATIONAL STUDENT OFFICE

P.O. Box 1214, UNIVERSITY OF TASMANIA, LAUNCESTON, 7250

March 15 1994

Dear International Student,

I am an international student at the University of Tasmania, Launceston and I hope that you will be able to assist in my research project in the Department of Education. Only 20 minutes of your time is needed to complete the enclosed questionnaire.

My research is concerned with finding out the opinions of international students about the International English Language Testing System (IELTS). If you have not taken the IELTS test, please return the questionnaire to me with only part 1 completed.

The information from your questionnaire will remain private, that is, confidential and anonymous (your name will not be recorded). The results are for my own postgraduate research and will not affect any IELTS test you take in the future. An information sheet is included to explain the purpose of the research to you.

The questionnaire is designed so that you can complete it easily. After completion, put it inside the envelope and place it in the box marked **IELTS Survey** outside the International Student Office in the Sir Raymond Ferrall Centre.

If you prefer to post your questionnaire, please write my name (Teresa Thiel) on the envelope and send it to the International Student Office. Participation in the research project is voluntary. Please sign one copy of the statement of informed consent if you are willing to be part of the survey of international students at the University of Tasmania. **Remember** to return the signed statement with your questionnaire.

You can contact me personally through the International Student Office or the English Language Centre if you have any questions or additional comments. Please return the questionnaire before April 1st. Your quick response and help are appreciated.

Yours sincerely,

Teresa Thiel
English Language Centre

Dr Thao Lê
Education Department
(Research Supervisor)

.....

Appendix V

INTERNATIONAL STUDENT OFFICE
P.O. Box 1214, UNIVERSITY OF TASMANIA, LAUNCESTON, 7250

INFORMATION SHEET

Title of investigation: International Students' Perceptions of IELTS

Chief investigator: Dr Thao Lê (Education Department)

Research student: Teresa Thiel

What is the purpose of the study?

This research is concerned with finding out the opinions of international students about the International English Language Testing System (IELTS), which is a language proficiency test taken by students entering Australian universities and colleges.

Why have you received this questionnaire?

The questionnaire has been sent to all international students at the University of Tasmania, Launceston campus.

What should you do?

Complete the questionnaire. This will take about 20 minutes. Place it in the envelope with one copy of the consent form signed by you. Return it to the box marked IELTS Survey outside the International Student Office. Return the questionnaire before April 1. You may post it if you prefer.

Is the information confidential?

The information from your questionnaire will remain private (confidential). Your name will not be recorded in the research report. Your opinions and comments will not affect any IELTS test you take in the future.

Is participation voluntary?

Participation in the research is completely voluntary. You may withdraw from the investigation at any time without any effect on your studies or on any future IELTS test.

Who can you contact if you have a question?

If you have any questions about the investigation, please contact Dr Thao Lê in the Education Department directly. If you have any difficulties with the questionnaire, please contact Teresa Thiel through the International Student Office.

Who can you contact if you have any worries?

If you have any personal concerns about participating in the investigation, you may contact the University Student Counsellor.

This study has been approved by the University of Tasmania Ethics Committee (Human Experimentation).

If you are interested in the results of the study, they will be available to you through the International Student Office in September 1994. Please keep this copy of the information sheet for your own reference. Return one copy of the consent form signed by you, and keep one copy for your own reference.

UNIVERSITY OF TASMANIA
International English Language Testing System (IELTS) Survey

PART 3

1. How many times have you taken the IELTS test in your own country?

☐ never ☐ once ☐ twice ☐ more than twice
2. How many times have you taken the IELTS test in Australia?

☐ never ☐ once ☐ twice ☐ more than twice
3. When was your last IELTS test? _____

(month) (year)
4. What is your overall IELTS band score?

☐ 4.0 or less
☐ 4.5-5.0
☐ 5.5-6.0
☐ 6.5-7.0
☐ more than 7.0
5. Have you taken a Test of English as a Foreign Language (TOEFL)?

☐ yes
☐ no (go to Part 4)
6. What is your TOEFL score?

☐ below 550
☐ 550-569
☐ 570-599
☐ above 600
7. When did you take the TOEFL test? _____

(month) (year)
8. If you have taken both the IELTS and the TOEFL, would you agree to be interviewed personally by the researcher?

☐ yes ☐ no

PART 4

1. To what extent does IELTS test any of the following?

grammar accuracy	<input type="checkbox"/> a lot	<input type="checkbox"/> some	<input type="checkbox"/> a little	<input type="checkbox"/> none
vocabulary (general)	<input type="checkbox"/> a lot	<input type="checkbox"/> some	<input type="checkbox"/> a little	<input type="checkbox"/> none
vocabulary (subject-specific)	<input type="checkbox"/> a lot	<input type="checkbox"/> some	<input type="checkbox"/> a little	<input type="checkbox"/> none
reading (general)	<input type="checkbox"/> a lot	<input type="checkbox"/> some	<input type="checkbox"/> a little	<input type="checkbox"/> none
reading (subject-specific)	<input type="checkbox"/> a lot	<input type="checkbox"/> some	<input type="checkbox"/> a little	<input type="checkbox"/> none
writing (general)	<input type="checkbox"/> a lot	<input type="checkbox"/> some	<input type="checkbox"/> a little	<input type="checkbox"/> none
writing (subject-specific)	<input type="checkbox"/> a lot	<input type="checkbox"/> some	<input type="checkbox"/> a little	<input type="checkbox"/> none
listening (general)	<input type="checkbox"/> a lot	<input type="checkbox"/> some	<input type="checkbox"/> a little	<input type="checkbox"/> none
listening (subject-specific)	<input type="checkbox"/> a lot	<input type="checkbox"/> some	<input type="checkbox"/> a little	<input type="checkbox"/> none
speaking	<input type="checkbox"/> a lot	<input type="checkbox"/> some	<input type="checkbox"/> a little	<input type="checkbox"/> none

Comments _____

2. Which module of the IELTS for reading and writing did you take?

- ☐ module A (Physical Sciences)
- ☐ module B (Biological Sciences)
- ☐ module C (Humanities and Social Sciences)
- ☐ General Training module

3. Did you have background knowledge in a specific subject for the module you chose?

☐ yes ☐ no

4. If **yes**, was your background knowledge helpful in the IELTS test?

☐ very helpful ☐ helpful ☐ not helpful ☐ not sure

Comments _____

PART 5

1. What is the level of English of an international student who scores band 6.0 on the IELTS test?

☐ excellent ☐ very good ☐ good ☐ fair ☐ limited

2. How would you describe your performance overall on the IELTS test?

☐ excellent ☐ very good ☐ good ☐ fair ☐ limited

Comments _____

3. How would you describe your level of English language when you took the IELTS test?

☐ excellent ☐ very good ☐ good ☐ fair ☐ limited

Comments _____

4. Is band 6.0 on the IELTS test a realistic entry level to study at a college/university in Australia?

a. for an undergraduate course ☐ too high ☐ realistic ☐ too low

b. for a postgraduate course ☐ too high ☐ realistic ☐ too low

Comments _____

5. Does the IELTS test predict an international student's ability to study in English?

☐ yes ☐ no

Comments _____

6. What other factors predict an international student's ability to study in English?

PART 6

1. Is there enough time given in the IELTS test to answer the questions in each section?

reading (55 mins)	<input type="checkbox"/> too much	<input type="checkbox"/> adequate	<input type="checkbox"/> too little
writing (45 mins)	<input type="checkbox"/> too much	<input type="checkbox"/> adequate	<input type="checkbox"/> too little
listening (30 mins)	<input type="checkbox"/> too much	<input type="checkbox"/> adequate	<input type="checkbox"/> too little
speaking (11-15 mins)	<input type="checkbox"/> too much	<input type="checkbox"/> adequate	<input type="checkbox"/> too little

Comments_____

2. What is the level of difficulty of each section of the IELTS test?

reading	<input type="checkbox"/> very difficult	<input type="checkbox"/> difficult	<input type="checkbox"/> fair	<input type="checkbox"/> easy
writing	<input type="checkbox"/> very difficult	<input type="checkbox"/> difficult	<input type="checkbox"/> fair	<input type="checkbox"/> easy
listening	<input type="checkbox"/> very difficult	<input type="checkbox"/> difficult	<input type="checkbox"/> fair	<input type="checkbox"/> easy
speaking	<input type="checkbox"/> very difficult	<input type="checkbox"/> difficult	<input type="checkbox"/> fair	<input type="checkbox"/> easy

Comments_____

3. Place your IELTS test results in order of your performance.

<input type="checkbox"/> reading	<input type="checkbox"/> writing	<input type="checkbox"/> listening	<input type="checkbox"/> speaking
----------------------------------	----------------------------------	------------------------------------	-----------------------------------

1 highest
2
3
4 lowest

PART 7

1. How important is each skill in your course of study?

reading	<input type="checkbox"/> very important	<input type="checkbox"/> important
	<input type="checkbox"/> not important	<input type="checkbox"/> not sure
writing	<input type="checkbox"/> very important	<input type="checkbox"/> important
	<input type="checkbox"/> not important	<input type="checkbox"/> not sure
listening	<input type="checkbox"/> very important	<input type="checkbox"/> important
	<input type="checkbox"/> not important	<input type="checkbox"/> not sure

speaking ☐ very important ☐ important
 ☐ not important ☐ not sure

Comments _____

2. Are the reading and writing tasks in the IELTS test like the tasks given in your course of study?

reading ☐ the same ☐ similar ☐ dissimilar
writing ☐ the same ☐ similar ☐ dissimilar

Comments _____

Thank you for your time and cooperation. When completed, please put this questionnaire and the consent form inside the envelope. Place the envelope in the box labelled IELTS SURVEY outside the International Student Office.

Appendix VII

UNIVERSITY OF TASMANIA, LAUNCESTON

IELTS INTERVIEW

Date _____

SECTION A - BACKGROUND DATA

1. Nationality _____

2. First language _____

3. undergraduate () postgraduate ()

4. Title of Degree course

5. How many times did you take the **IELTS** test? _____
(International English Language Testing System)

What was your overall band score? _____

When was your last IELTS test? _____
(month) (year)

6. How many times did you take the **TOEFL** test? _____
(Test of English as a Foreign Language)

What was your score? _____

When was your last TOEFL test? _____
(month) (year)

SECTION B - TARGET SITUATION

1. What English language skills do international students need in order to study successfully at this university?
2. Do international students need different language skills for undergraduate and postgraduate studies?

yes no not sure

Comments _____

3. Which is the most important English language skill needed by an international student?
4. What are the reasons for its importance?
5. In your opinion, which language skills are most important in your studies at this university?

SECTION C - TEST EXPERIENCE

1. The IELTS and the TOEFL are tests of English language proficiency.
What is your understanding of 'English language proficiency'?
2. Do you think the IELTS tests English for academic purposes?

yes no not sure

What are the reasons for your opinion?

3. Do you think the TOEFL tests English for academic purposes?

yes no not sure

What are the reasons for your opinion?

4. Are there any differences between the TOEFL and the IELTS tests?

yes **no** **not sure**

5. If **yes**, what are these differences?

6. Does the IELTS test a student's background knowledge in specific subjects?

yes **no** **not sure**

Comments _____

7. Does the TOEFL test a student's background knowledge in specific subjects?

yes **no** **not sure**

Comments _____

SECTION D - TEST METHODS

1. Is it possible to prepare for the IELTS test?

yes **no** **not sure**

What language activities help to prepare a student for the test?

2. Is it possible to prepare for the TOEFL test?

yes **no** **not sure**

What language activities help to prepare a student for the test?

3. What are the main test methods used in the TOEFL?

4. What are the main test methods used in the IELTS?

5. Is the IELTS test academically oriented, that is, directly connected to academic study?

yes

no

not sure

6. Which parts of the IELTS test do you think are most like real-life, that is, authentic?

7. Is the TOEFL test academically oriented, that is, directly connected to academic study?

yes

no

not sure

8. Which parts of the TOEFL test do you think are most like real-life, that is, authentic?

Appendix VIII

INTERNATIONAL ENGLISH LANGUAGE TESTING SYSTEM

INTERVIEW TAPESCRIPT

Target Situation

- R From your experience here as international students, what English language skills do you need to study successfully at this university?
- S1 Basically, reading and writing are most important at first place.
- S2 For students who will study in coursework I think beside reading and writing skills, important thing also is listening because and the way, how the lecturers speak in certain countries should teach the use because like the first time I came here I found the Australian dialect very accent so it was very difficult to catch.
- R Anybody else?
- S3 I think all of them are important
- S1 Yes, my first experience here was terrible. I found that listening, it is too difficult to understand what the lecturer talk about. Then I have to control this weakness by reading and for doing the assignment writing, that is the most important.
- S4 For my experience, among the three, reading, listening and writing, everything is most important.
- R Do you think international students need different language skills for undergraduate and postgraduate studies?
- S2 Yes I think it's really need for that, because, for example, in postgraduate studies the students are required to perform in the paper or the best way to write, for example in literature review, I never have experience before in university making review like this and so I think it is really important because actually we have to know exactly how to cope, how to cipher information from the book and how to write it in the paper as a literature review.

- R Okay, it might be different for different courses? Anybody else?
- S5 I think for postgraduate study, especially writing.
- R So more the emphasis on writing perhaps at postgraduate level. Okay.
- R What about you personally now, which skill is most important for you?
- S3 For my studies?
- R Yes, for your studies. You said all of them.
- S3 Yes, but I think the most important for me is speaking because we have a lot of discussion in class but I have this problem so I always keep quiet.
- R So that's the one you identify. Yes?
- S1 If we want to participate in discussion in class, speaking and listening is all important.
- S4 It depends on the time.
- S2 For masters students in aquaculture this year we have to do graduate diploma course, coursework one year. In this year the important skill that is needed is notetaking and listening and of course writing some assignments as I mentioned already. Last year, second year, we really need skill of writing because in this last year we have to spend our time in writing manuscript presentation.
- S4 Thesis.
- S2 Thesis.
- S4 And work experience report.
- S2 And work experience report. Actually it's easier to get the information from our fieldwork but some Australian students even finds it very difficult to write down.
- R So you're saying it changes as you go through your course, the importance for each skill changes?

S4 One of the reasons, in aquaculture every week we go to field asking for information from the farmers and it's very difficult to catch what they say and their pronunciation is quite different and it's very hard to catch and that's very important for us in aquaculture first year.

Test Experience

R Let's go on to the next section which is about test experience. The IELTS and the TOEFL are tests of English language proficiency. What do you understand by English language proficiency? What do you think ...

S3 The meaning?

R that means? Yes.

S3 The ability to perform in English.

S4 I think English proficiency here means our ability to do courses in universities including reading, writing, and speaking, listening, all things especially in academic aspects.

R Do you see proficiency as something divided into skills: reading, writing?

S1 No

S5 Includes everything

R Includes everything?

S1 Overall

R So are we happy with ... (S3) definition? Does everybody agree, ability to perform in English?

All Yes

R Do you think the IELTS tests English for academic purposes?

- S1 Yes.
- S3 Absolutely.
- S5 Absolutely.
- S2 For me not really because, for example, in IELTS test not one types of IELTS test, there are three type. Right? A, B, C, ...
- S1 There is four actually.
- S2 A for Physics.
- S1 Modules.
- S2 Modules. A for Physics, B for Biology Science, C for Social Science. For example, in Biology Science I come from fisheries, aquaculture, sometimes the test is come from Medical Science. I don't think that's really good perform our ability in my academic background, so it is not really in that sense English, academic English.
- R What does it represent then if it's not academic English? You said absolutely academic English. Yes?
- S1 I said it's absolutely academic purpose because, yes, that problem perhaps what ... (S2) experienced never happened to me so I also feel that those four skills, reading, writing, listening and...
- R Speaking
- S1 Speaking, they test all, that is the most appropriate field that need to be tested.
- S4 The answer I think is this, probably ... (S2) answers number six. I agree with ... (S2).
- R Well, let's go to number six because that's connected. Does it test your background knowledge in specific subjects?
- S4 Maybe not.
- S1 Not always.

- S2 Not always.
- S4 My experience always, exactly the same as ... (S2).
- S3 But I think it is just not important because the IELTS test the reading that sometimes not relevant to your background, but just for test how you can understand only the basic English. I don't think it will go through your deep knowledge or something.
- S2 Yes I agree, but I found my experience one time when the IELTS test deal with my field with my ...
- S4 Knowledge background.
- S3 My knowledge background I will get the best mark, pass.
- S1 So lucky in the reading.
- R Maybe a matter of luck?
- S1 They provide the background reading before to help you when you got to test the background so that if you're not familiar with the background reading before you proceed in your test.
- R Can you explain what you mean?
- S1 I mean anybody they don't understand the field can be guided by the test because of their previous reading, because of their reading at the previous level. Same field there are several readings before you proceed with the writing in special field. You are given the reading material, the reading regarding that field.
- R What do you mean given it, given it by whom?
- S1 Given it by the exam.
- R You mean in the exam paper you have the reading passages?
- S1 Yes. I think it guides you.
- R I see, to help with the writing you mean?

- S1 Yes.
- R What about the TOEFL test, you've all taken the TOEFL test, does that test English for academic purposes?
- S5 There are three parts in TOEFL, listening, grammar, reading and reading topic is different from IELTS, but I think TOEFL is academic purpose.
- S3 Yes, I agree.
- S1 It seems that TOEFL tests your comprehensive knowledge not academic.
- R What do you mean comprehensive knowledge?
- S1 Your vocabulary, your grammar.
- S2 I'm not sure. Probably I agree with ... (S1). TOEFL tests for studying. This test also applicable for someone who just go abroad. This test only assess English, but not academic purpose.
- R So your background knowledge, is that helpful in the TOEFL test?
- S4 It is general
- S1 For me it is general.
- R From what you've said, you can see some differences between TOEFL and the IELTS test? Do you think there are some differences between them?
- S3 Yes.
- S1 Yes.
- S4 The first one I think we have discussed, IELTS test more disposed to the academic purpose rather than the TOEFL, it's general. The most important thing I think TOEFL is more objective, while IELTS test is not. What I mean is if TOEFL saw our answer is really objective so no-one ..., but for IELTS test it depends on the person, like speaking. For example, if you test somebody maybe you can pass the test with same person you gain higher mark. That's our experience, so we say that's a

good teacher. She will give good mark, and another thing is the materials, sometimes the materials they use for last month some of the materials they use for the next or the next year.

R Which test do you mean?

S4 IELTS test. Yes. TOEFL never. In my experience took TOEFL four times all different, but IELTS test.

R Have any of you who have taken the IELTS more than once had the same experience, the same test material?

S3 I took only once.

R Are there any other differences?

S5 Yes, I think TOEFL has grammar section. IELTS test does not have grammar section so it is really different. Also TOEFL each question is independent, but IELTS test dependent question. It is different I think. IELTS is severe, but TOEFL not severe I think.

S3 One difference is that TOEFL test depends on the student the people who test. The score, it depends on how high the average score or something like that.

S5 IELTS independent.

S3 Could you understand me? The score depend on the group.

R Yes, all the results on TOEFL are compared with ..., all other students. Whereas, as you mentioned, with IELTS your score is ...

S5 Independent.

S4 My feeling is those people who are taking English course for IELTS test, they don't have as good grammatical knowledge as those people who are taking course for TOEFL. That's my experience. What I found here, our friends, for example, they just did IELTS test. Their grammar is not same as people experience TOEFL because in the IELTS test they don't focus or don't stress on the grammatical knowledge.

Test Methods

- R You've identified a lot of differences. Let's go to the last section which is test methods. Is it possible to prepare for the IELTS test?
- S2 Yes, of course.
- R What sort of language activities help to prepare you for the test, do you think?
- S2 What do you mean?
- R What type of preparation, what sort of classes or courses, what sort of preparation do you need? For example, can you just study past papers, or can you, do you need to follow a particular English course, or do you need to study independently?
- S2 I think first time really need an English course for a couple of months. Of course it depends on the background, I mean the ability of the student because student who has English ability poor maybe they need longer in English class than for student has better English background. Then, of course, we still need to read past tests because in my experience I got 550 TOEFL, then first time I try to fill the exercise IELTS test I got lower mark than other students who has already experienced in filling the IELTS test. Then after a couple of weeks I find myself much better than them because I think my background English is already better. Then I only need the skill how to fill this test.
- R So you mean there's a skill, a strategy for the test.
- S2 Right.
- S1 I think interaction, by simulating interaction with others, reading and discussion. Independent study of course more and more reading, and writing skill.
- R Anybody else?
- S3 One problem to prepare the student for the test is to give the strategy to manage the time to take the test.
- R Time?

- S3 Time manage...
- S1 Time management.
- R Time management. Is it possible to prepare for the TOEFL test?
- All Yes.
- R What sort of language activities do you think help prepare for the TOEFL?
- S1 Just study I don't think that you need special, because possible you just reading and grammar.
- S2 For listening we still need lecturer to help us, because sometimes we find the TOEFL test they speak too fast, and they pronounce strange to our ear. Sometimes we need tutorial or lecture to catch what the lecturer reading or what is actually spoken in the test. Sometimes native speaker couldn't catch the real words in the TOEFL test.
- S4 In the TOEFL test strategy also is very important, especially to manage time because it's very hard, and the reason why, if I compare with IELTS test, is in TOEFL we should interpret what the sentence means because there are four many choices and almost similar, so it's very hard, but in the IELTS test we just fill in the gaps with phrases so we need skill also for TOEFL actually.
- S3 For guessing.
- S4 Guessing, if you want to guess. Yes, also we can guess. We can get a good mark on this guessing, but for IELTS test if you don't know it you cannot do.
- S3 Especially in the vocabulary part you can guess.
- S4 In my experience in IELTS test sometimes we cannot catch what they say but because of the phrase is our background knowledge, sometimes we can fill it, this one maybe.
- S2 Yes that's right.
- S4 Yes, possible if that's our background, our knowledge.
- R Are you referring to the listening part in IELTS?

- S4 Listening, yes, we talk about listening IELTS.
- S5 Grammar section is very easy to get high score in TOEFL, if you do exercises, lots of exercises.
- R Lots of practice.
- S5 Lots of practice for grammar is very easy to get higher score.
- R So from what you've said you can see different test methods used between the IELTS and the TOEFL. So if you just wanted to summarize, what would you say the main differences are in the test methods used between IELTS and TOEFL.
- S3 Multiple choice is main difference because IELTS we just fill the form, but sometimes we have multiple-choice, but hardly I think. The TOEFL all are multiple-choice. The problem is that some students can guess.
- S5 Yes.
- S3 And the probability is higher than IELTS.
- S4 Vocabulary you mean?
- S1 Yes.
- R Do you think the TOEFL tests writing?
- S1 Sometimes.
- S4 In year maybe eight times TOEFL test, and four with writing.
- S5 TWE.
- R TWE, yes, Test of Written English. Did you take the test?
- S5 Yes.
- R You all took the Test of Written English?

- All Yes.
- S1 Speaking no.
- S3 No speaking.
- R What do you think about that? Do you think it 's necessary to have a speaking test?
- S1 It's impossible.
- S4 They want TOEFL to be really objective. That's the main difference between TOEFL and IELTS test.
- S2 Sometimes for academic purpose I think speaking is not important. Especially for masters degree with research the speaker is the supervisor, maybe no need to go to classroom to discuss, so it depends.
- R Is the IELTS test academically orientated, that means directly connected to academic study?
- S2 Yes.
- S4 Yes.
- S5 Directly.
- R Which parts of the IELTS test do you think are most like real-life, authentic, like the situation in your academic studies?
- S4 Situation in academic studies?
- R Which parts of the test do you think are like what you have to do at the university in real-life?
- S2 Writing, how to write an essay, that's similar to university, what else, I think ...
- R Just writing is most like the real situation?
- S3 I don't think no any part that can relate to my course. It's very different.

- R Would you like to explain in what way it's different?
- S3 For instance, writing, yes, sometimes it's a bit, use some skill, for instance, introduction, body, and conclusion, but in my real course, my assignment, then you can say it's different from the IELTS test, and listening, listening is just the general situation, but in my class we talk about business, management and some vocabulary that I never know before.
- S1 I think reading and writing is like real situation.
- S4 I can compare to experience. In IELTS test we practice how to describe a process. That is exactly like our course, process, and that's like in IELTS test.
- R And finally the TOEFL test? Academically orientated, connected to your course of study at all?
- S3 No.
- S4 General.
- S1 Too general.
- R Too general, what do you mean?
- S1 Yes, it's vocabulary. It seems that this one only test grammar and seems to me TOEFL tests your overall, except speaking, language ability in a very lucky situation. You can guess the answer because it's multiple-choice.
- R Anything else you would like to comment on?
- S5 I think many chances to take TOEFL. IELTS test not many chance, and I can't take it during three months.
- R That's a regulation of IELTS to leave it three months before you take another test.
- S5 IELTS expensive.
- R In Japan, the TOEFL is more recognized?

S5 Yes.

S3 The same in Thailand, the TOEFL test, with the location for the test there are more than IELTS. IELTS test just only in Bangkok, but TOEFL in Chiang Mai or in south part of Thailand you can take the test. IELTS no.

S4 IELTS test isn't very popular.

S3 Not popular.

S4 Less popular I think in every country.

R Why do you think it's less popular?

S4 Because maybe it's a new one, very new.

S2 Also only in certain country the IELTS result accepted. If we are going to America maybe they don't take IELTS test.

S4 But TOEFL can be accepted for all countries.

S2 Even in Australia.

S4 Australia also.

S3 When I applied I used the TOEFL test.

R Anything else?
Then, thank you very much for your time.

THE END