CHAPTER ONE
INTRODUCTION

Applied Linguistics and Language Testing

Applied linguistics cannot be defined easily because of its interdisciplinary nature. According to Davies et al. (1999, p. 9), there are two opposing views of applied linguistics as a discipline: the weak view and the strong view. The weak view of applied linguistics is explained by the traditional definition of applied linguistics as the application of linguistic theories and procedures to other disciplines, e.g. the study of literary texts and writing syntax for a computer program. This version of applied linguistics has been called ‘linguistics applied’ as opposed to the strong version, ‘applied linguistics’.

Applied linguistics, the strong version, is “a practice-driven discipline that addresses language-based problems in the real-world contexts” (Grape, 2002, p. 10). As this definition suggests, applied linguistics focuses on trying to resolve language-based problems that people encounter in the real world, especially those of language education (first-language, second-language and foreign language teaching and learning). The main distinguishing characteristic of applied linguistics is its concern with professional activities whose aim is to solve real-world language-based problems, which means that research in this area touches on a particularly wide range of issues – psychological, pedagogical, social, political and economic as well as linguistic. As a consequence, applied linguistics research tends to be interdisciplinary.

It is generally agreed that applied linguistics is not simply the application of research done in linguistics. On the one hand, applied linguistics has a look beyond linguistics for relevant research and theory, so applied linguistic research often involves the synthesis of research from a variety of disciplines, including linguistics. On the other hand, applied linguistics has been responsible for the development of original research in a number of areas of linguistics, e.g. bilingualism, literacy, etc.

Applied linguistics, being mostly concerned with language teaching and learning, makes use of language testing as a tool for solving the language teaching and learning problems. Language testing is concerned with major problems in language use and is firmly committed to the measurement of language teaching and learning. Because of its psychometric techniques which make it prove itself and its success in doing what it claims to do, language testing has become increasingly
important and occupied a central position within applied linguistics. It assesses the extent to which our language learning has achieved its set goals. In the words of Davies (1990), language testing offers

a methodology for carrying out empirical research in applied linguistics, whether that research is language testing (tout court) research; investigating language acquisition, judgments, intelligibility studies, comprehension and use; or comparative experiments in language teaching methodologies and materials. (p. 2 [original italics])

Developments in Language Testing

Language testing during the last thirty years has moved itself from having only a practical status within applied linguistic research to having both a practical and a theoretical status. According to Bachman (1991), advances in language testing have occurred in three areas: theory, methodology and test development. Advances in the first area relate to “the development of a theoretical view that considers language ability to be multi-componential and recognize the influence of the test method and test taker characteristics on test performance”. Advances in the second area relate to “applications of more sophisticated measurement and statistical tools” while advances in the third area relate to “the development of “communicative” language tests” (p. 671). Advances in these three areas are detailed below.

Theoretical Advances

One of the major preoccupations of language testers in the past three decades has been investigating the nature of language proficiency. In 1980 the “Unitary Competence Hypothesis” (Oller, 1979), which claimed that language proficiency consists of a single global ability, was widely accepted. However, this view was challenged by empirical studies and was later replaced by the view that language proficiency is multi-componential, consisting of a number of interrelated specific abilities as well as a general ability. The model of language test performance proposed by Bachman (1990b) represents progress in this area because it includes both components of language ability and characteristics of test methods, thereby making it possible to make statements about actual performance as well as underlying abilities. Later, this model was developed and eventually superseded by Bachman and Palmer (1996). Further progress was conducted in the application of theoretical models of language proficiency to the design and development of language tests (e.g. the

Another theoretical advancement of the theory relates to the understanding of the effects of the method of testing on test performance. A number of empirical studies conducted in the 1980s and 1990s clearly demonstrated that the kind of test tasks used, just like the abilities to be measured, can affect test performance (e.g. Shohamy, 1984, cited in Bachman, 1991; Riley & Lee, 1996, cited in Bachman, 2000). There have also been important developments in the refinement of approaches to the analysis of test method characteristics, of which content is a substantial component, and the investigation of how specific characteristics of test method affect test performance (e.g. Bachman, Davidson & Milanovic, 1996; Carr, 2006). In addition, more focus has been given to the role of raters in the assessment process. Such factors as the rater background, native language and previous rating experience have been found to effect differences in the rater behavior (e.g. McNamara, 1996).

A third area of progress occurs in research into the characteristics of test takers and how these are related to test performance. Kunnan (1998a, cited in Bachman, 2000) lists over twenty studies investigating the characteristics of test takers that affect their performance on the test. These characteristics are academic background, native language, culture, gender, and field dependence. Such studies demonstrate the effect of various test taker characteristics on test performance. Furthermore, researchers such as Bachman and Palmer, in their model of language ability (1996), assert that such characteristics must be considered in both the design of language tests and the interpretation of test scores.

**Methodological Advances**

Many of the advances mentioned in the previous section have been facilitated by advances in the tools available for test analysis. According to Bachman (1990), these advances “have been in three areas: psychometrics, statistical analysis, and qualitative approaches to the description of test performance” (p. 676).

Research in the area of language testing has used psychometric tools to analyze test items in terms of their difficulty and discrimination. Before the 1980s, classical item analysis (CIA) was the only psychometric tool used for analyzing test items. However, this kind of analysis has its own problems. One of the problems is that “the item and score statistics we obtain from tests developed with classical IA are essentially sample-based descriptive statistics” (Bachman, 2004, p. 139). This means
that the information obtained from this analysis is always limited to the sample under test.

The 1980s saw the applicability of more sophisticated psychometric tools such as Item Response Theory (IRT), and Generalizability Theory (G-theory). The greater advantage of the newer techniques of IRT is that they are sample-free. The techniques of IRT, which have become the dominant test development methodology for large-scale standardized language proficiency tests, are capable of making strong predictions about the characteristics of individual test items, individuals’ performance in individual test items and their levels of ability. However, both CIA and IRT can be used to identify items that fail to function properly and to select the items that have the characteristics needed. The third psychometric tool, the G-theory, is used to specify and estimate the relative effects of different factors on test scores. G-theory is “useful in such areas of test development and research as estimating the effect caused by varying the number of items or tasks in a test, or the number of raters involved in scoring each performance” (Davies, et al. 1999, p. 67).

With the expansion of distance learning, the development of the internet, and the increasing availability and accessibility of both software and hardware, computers have become widely used in education. The use of computer technology in language assessment has been growing out of the need to measure English communicative skills with maximal accuracy and efficiency. Language testers have recently succeeded in developing computer/web-based language tests which are more authentic and interactive than their paper and pencil counterparts. Among these are the TOEFL CBT and more recently the TOEFL iBT developed by Educational Testing Service (ETS).

A major methodological advance in the area of statistical analysis has been the use of confirmatory factor analysis which is used as a productive tool in the process of construct validation. Another methodological development within the area of statistical analysis is the application of structural equation modeling to language testing research. Structural equation modeling (SEM) extends factor analysis. It has facilitated the investigation of the “factor structure of the measures we use and the relationships among these factors, or latent variables” (Bachman, 2000, p. 6).

The third methodological advance is the use of qualitative approaches in language testing research to investigate a number of research questions. These questions, according to Bachman (2000), include “the effects on test performance of
test takers’ characteristics, the processes and strategies they use to respond to assessment tasks and the characteristics of assessment tasks themselves” (p. 6). There are different qualitative approaches to language testing research, which include “expert judgments, introspective and retrospective verbal reports, observations, questionnaires and interviews, as well as text analysis, conversational analysis and discourse analysis” (p. 7).

It is well-known now that a language test score cannot be interpreted simply as an indicator of the language ability we want to measure; it is also affected to some extent by the characteristics of the test taker, and the strategies the test taker employs in attempting to complete the test task. What makes the interpretation of the test scores particularly difficult is that these factors interact with each other. The particular strategy adopted by a given test taker, for example, is likely to be a function of both the characteristics of the test task and the test taker’s personal characteristics. At the same time, the knowledge of the complexity of language test performance, along with the methodological tools available, provide a basis for designing and developing language tests that are potentially more suitable for specific groups of test takers and more useful for their intended purposes.

**Advances in Language Test Development**

In the 1960s and 1970s, language testing practice was informed by the ‘discrete point’ approach to test design. This practice was reflected in large-scale language tests (e.g. the TOEFL test) and in the literature on language testing at that time (Lado, 1961; Harris, 1969). In the 1980s, the work of applied linguists such as Michael Canale and Merrill Swain (1980) proved extremely influential and fruitful to language testing. Bachman (2000) reports that Canale and Swain viewed language use as “the creation of discourse, or the situated negotiation of meaning, and of language ability as multicomponentail and dynamic” and that this view of language use

… forced language testers out of their narrow conception of language ability as an isolated ‘trait’, and required them to take into consideration the discoursal and sociolinguistic aspects of language use, as well as the context in which it takes place. (p. 3)

Bachman also observes that Morrow’s (1979) ‘provocative’ paper, which proposed a ‘communicative’ approach to language testing, set off a heated debate, which persists to this day, about the nature of communicative, or authentic, language tests. A full
description of these views or approaches to language test development is given in detail in Chapter Two.

These developments in language testing have taken it away from the center of applied linguistics to a separate existence “thus creating, or threatening to create, a new discipline” (Davies, 1990, p.71). This move of language testing from subject to discipline within applied linguistics has also meant more research and more theory which do not appear to be of direct relevance to language teaching. Language testing has extended the scope of testing from being traditionally concerned with the measurement of the learner’s output (which aims at providing a clear picture of the test taker’s language ability and attainment) to being concerned with the evaluation of courses, materials, and projects using both quantitative and qualitative measures (which aims at providing as full a picture as possible of the project, course or materials in order to prove the adequacy or otherwise of that project, course or materials).

The present study attempts these two types of evaluation. It attempts an evaluation of the 2005 admission test given to Yemeni secondary school leavers intending to join the English departments at Taiz University in particular and of the admission testing program in general. After this evaluation is made, an admission test will be constructed to measure the test takers’ language ability. The first type of evaluation, the evaluation of the test and of the testing program, will be made using qualitative and quantitative research instruments. The construction of the test, on the other hand, would be based on Bachman and Palmer’s model of communicative ability (1996) and would necessarily meet the current developments in the theory and practice of language testing, especially in communicative language testing.

Background of the Study

Yemen: The Country

Yemen, officially The Republic of Yemen, is a Middle Eastern country with an area of 530,000 square kilometers. Its is located in the Arabian peninsula, bordering the Arabian Sea and the Gulf of Aden to the south, the Red Sea to the west, Oman to the east and Saudi Arabia to the north. Yemen is divided into 19 governorates: Abyan, Aden, Ad Dali, Al Bayda, Al Hudaydah, Al Jawf, Al Maharah, Al Mahwit, Amran, Dhamar, Hadramawt, Hajjah, Ibb, Lahj, Marib, Sa’dah, Sana’a, Shabwah and Taiz. Sana’a is the capital and the largest city in Yemen.
According to the world factbook (2007), the total population of Yemen is 21,456,188 (July 2006 est.). Almost 99.94% of the total population are Muslims. The rest are Jews, Christians and Hindus. Arabic is the official language in Yemen and English is the only foreign language taught in all Yemeni schools and universities.

Out of the total population of Yemen, only 50.5% are literate, and the majority of these are males. The reason for this low level of literacy in Yemen is that Yemen was ruled by the Imam (monarch) dynasty which kept the Yemeni people in long closure and complete isolation from the outside world, especially the non-Arab world. Fortunately, this ruthless monarchy system was overthrown in 1962 by the republican revolution. Since then, the country has witnessed radical changes at all levels, especially education. One of the main changes in education has been the establishment of a formal education system and the inclusion of English as a subject in the school and university curricula. The following part of this chapter deals with English education in Yemen and diagnoses its strengths and weaknesses with regard to the curriculum, teaching methods, textbooks, teachers and students.

An Overview of English Education in Yemen

Education is important for economic growth and social development, and English education is important for successful integration into the global world. Being aware of the importance of education, successive Yemeni governments have since independence been working on improving the status of education generally and English education in particular. English is now an obligatory subject in the curricula of Yemeni schools and universities. School students take English classes from the seventh grade onwards, and university students take two to four English classes in their eight semesters of university education. There have also been constant attempts to improve infrastructure, streamline English curricula, qualify Yemeni youths as English teachers, and train these teachers in the latest teaching methodologies.

Despite these efforts, there is still a general dissatisfaction with English education on the part of all involved (i.e. students, parents, teachers and government officials). The Education budget has increased, the number of educational institutions increased, the number of students increased but the educational standards and output do not show any significant improvement. The quality of education at both levels, i.e. school and university, is generally below satisfactory and many of the graduates can hardly write their names in English.
There have also been several attempts to rectify the relative unsuccess of English education in Yemen. The last fifteen years have witnessed the establishment of a number of private institutions for teaching English. Private schools have introduced English classes right from the first grade and, in some schools, used English as the medium of instruction. Private universities have also been established and supposedly offer better English programs. Graduates from these schools and universities, however, are not always better performers in English. But this private education is itself a luxury not many parents can afford. The less expensive private language institutes offer a relatively more affordable option. These institutes offer English courses to students of all levels. The duration of each course of instruction is four to six weeks and the course fee is fifteen to thirty dollars. The problem, however, is that parents believe in short-terms results and when they see that their children have not learnt much for the money paid, they discontinue private education. The educational scenario is obviously complicated.

Clearly, there is recognition on the part of parents and the government of the increasing importance of English. Two weekly newspapers (The Yemen Times and The Yemen Observer) are printed in English and the Yemeni local TV channel broadcasts an abridged version of the daily news in English. The English education scene in Yemen, however, is not encouraging and needs more resources and even more efficient management. The next two sections of this chapter will attempt an outline of the curricula, teaching methods, teachers, etc. with reference to the teaching of English at the Yemeni government public schools and government universities. This will be followed by an overview of the English teaching situation at the two departments of English at Taiz University.

**English education at the school level.** One of the main developments in the educational system in Yemen has been the inclusion of English as a subject in the school curriculum. English is presently introduced to the seventh graders at government schools and it remains with them until they graduate from secondary school. There is a government plan to start English classes at the fourth grade but this plan has not come into effect yet.

There have been other attempts to improve the teaching and learning of English at Yemeni schools. A major concern has been the development of English textbooks which could satisfy the aims of teaching English in Yemen and prepare the students for the world outside school. The ministry of education, in collaboration with
the British council, designed the EFY (English for Yemen) series in 1975. This series was designed to develop competence in the structural and communicative aspects of English. The teachers, however, were not trained in the communicative approach to language teaching and so they focused solely on the structural aspects of language, especially grammar and reading. The EFY series remained in use in Yemeni schools until it was replaced by the Crescent English for Yemen (CEY) later in 1995, a series of textbooks published by the Oxford University Press. The CEY was designed in accordance with the communicative language teaching/learning and aims at the development of the Yemeni learner’s productive and receptive skills in English. The series is still taught at Yemeni schools at the time of writing this dissertation.

In spite of all these attempts to develop a ‘communicative’ syllabus, the teaching of English at Yemeni government schools remains largely ‘traditional’. There are a number of reasons for this. Firstly, the teachers receive little training in the communicative approach to language teaching. Secondly, the classes are overcrowded and the class duration is only 45 minutes. Individual attention to students is ruled out. Thirdly, the teaching aids needed in communicative classrooms are largely unavailable to teachers. These reasons combine to dictate the use of the grammar-translation method, regardless of the aims of the textbook writers. Instead of giving individual attention to the productive and receptive language skills of around a hundred students, teachers shift their attention to the more manageable rules of the language. Thus, the English classroom becomes a lecture on grammar and translation. Another consequence of the reliance on the grammar-translation method is the teacher-centeredness of the classrooms. All the language work is done by the teacher and the students’ role is to sit quietly and receive passively.

The complications of teaching English at Yemeni schools do not stop here. Most school teachers are under-qualified. Their knowledge of methodology and assessment is basic and they do not attempt to even update this basic knowledge. Besides, their English language proficiency is itself substandard. These problems are traceable to the teachers’ own preparation at the university level. In some cases, the students who join the English department at the Faculty of Education are not really the best candidates. The problem here is one of selection, caused by the use of inappropriate selection tools. In other cases, it is the course content and teaching methodology at the college level which disadvantages the would-be teachers. The teaching at the university level is itself teacher-controlled and there is little room for
discussion and interaction (the problems at the university level are detailed in the next section). In either case, however, the EFL scenario at Yemeni schools is further complicated.

The students are further disadvantaged by the unofficial status of English in Yemen. The only language used in Yemen is Arabic and the use of English is restricted to the English classrooms and to a few other formal contexts. The students do not speak much English in the classroom and do not speak English at all outside the classroom. When these students join university to major in English, their proficiency is low and they cannot cope with the requirements of university education.

*English education at the university level.* The university input is clearly substandard. When school leavers join the English departments, their English proficiency is low and they need much help. The teaching program at the university, however, cannot be adapted to the level of proficiency of the entrants. Instead, it assumes a certain level of proficiency and builds on it. As a result, many of the entrants find themselves disadvantaged once again. When these students cannot cope with the courses offered at university, they resort to examination-passing techniques which help them move through the semesters but which do not help them improve their language proficiency. Besides, the universities share some of the school problems. The classrooms are over-crowded and teaching aids not available. It is a vicious circle. A certain level of proficiency is pre-supposed by university but which was not promoted by school. When these students graduate from university and become teachers, a certain level of proficiency is expected of them but which again was not promoted by university. The education scenario at the university level is also complicated. The rest of this section offers a closer look at one Yemeni university, namely, the University of Taiz, to which the researcher is affiliated and in which the present study is located.

Taiz University, the third largest university in Yemen, was established in 1993. Prior to that, Yemen had only two universities, Sana’a University and Aden University. The city of Taiz had three colleges affiliated to Sana’a University, viz. the college of arts, the college of education and the college of science.

Taiz University offers two four-year undergraduate programs in English, one at the Faculty of Arts and the other at the Faculty of Education. The English
department at the Faculty of Arts offers courses on English language, linguistics, translation and literature (see Appendix A for a detailed description of the program content). The English department in the Faculty of Education, on the other hand, offers fewer linguistics and literature courses and more language teaching courses, in English and in Arabic (program content is also offered in Appendix A). This difference in course concentration may be explained by the ‘goals’ of the two departments. Graduates from the Faculty of Arts usually end up in positions which require good knowledge of English. Whether they work as translators, office managers or newspaper editors, these graduates are expected to demonstrate good command of spoken and written English, a goal which is supposedly met by the introduction of extra linguistics and literature classes. Graduates from the Faculty of Education, on the other hand, take teaching positions. The department has to equip them with the theoretical and practical knowledge necessary for them to carry out their teaching duties successfully. Hence the concentration of courses on methodology and learner psychology.

An Overview of English Educational Assessment in Yemen

English educational assessment does not get due attention in Yemen. At the university level, very few departments offer courses on language assessment. The ministry of education does not offer training to in-service teachers on language evaluation and assessment. Moreover, there is little empirical research on teachers’ assessment practices and on program evaluation. The following synoptic overview will focus on educational assessment at the school and university levels.

English educational assessment at the school level. Yemeni school teachers receive little instruction on language assessment in their four-year course of study at the university level. When they take up their teaching positions, they also receive little on-the-job training on assessment and evaluation. As such, they can only depend on their experience as teachers and as students in the development of achievement tests. These tests are always teacher-developed tests. The content of these tests is a clear reflection of the teaching methods used by the teachers. The tests concentrate on the students’ grammatical ability and reading comprehension. There are a few attempts to include communicative activities in the form of written dialogues but these few individual attempts do not constitute a significant move towards a global concern with communicative language testing.
These teacher-developed tests are also always scored by the teachers themselves. The maximum mark for each test is 100 and the pass mark is 50. The students’ scores on these tests are used to make decisions on their language achievement. The teachers, however, have little knowledge of statistical measures related to test analysis. The students’ test scores are therefore not put to any kind of statistical analysis. The reliability of the test scores and the validity of the decisions based on them remain uninvestigated.

There are also virtually no standardized tests at any level. This may be explained by the fact that English education in Yemen is relatively young (about 40 years old) and that funds and resources in Yemen are allocated to other important issues such as building new schools, expansion of existing schools and the streamlining of school curriculum. The only exception here is the National Secondary School Leaving Certificate. This is a set of seven to eight examination papers administered by the ministry of education for secondary school students in order for them to graduate from school and qualify for university education. All third-year secondary school students from all over Yemen take the same set of papers. The English paper, which is one of the papers the students have to take, is not without problems. It focuses on the formal aspects of the language rather than on the communicative ability of the learner. According to a report submitted by the University of Cambridge Local Examination Syndicate (1985, p. 13, cited in Al-Hamzi, 1999, p. 15) the English papers are “seriously restricted in the range of skills tested, concentrating mainly on the factual recall of textbook information and sometimes encouraging the repetition of learned model answers based on text book exercises”.

**English educational assessment at the university level.** Achievement tests at the university level are also teacher-developed. These tests are also scored by the teachers themselves, who also have little knowledge of statistics and of the means by which test scores can be investigated for their reliability. In order for these tests to be more reliable and more efficient measures of language achievement, teachers need to be given training in language test construction and in test scores analysis and interpretation. The rest of this section, and this dissertation, will, however, focus on proficiency tests at the university level, namely, admission tests. Below is a brief overview of the admission testing situation at Taiz University.
Before getting officially admitted into the two departments of English at Taiz University, candidates have to take an admission test. Before 1998, the two departments at the faculties of Arts and Education constructed and administered separate tests to their applicants. In the last ten year, however, the two departments have used the same test. The test is constructed by one of the teaching staff at any of the two departments, preferably a Yemeni with a PhD. The test writer is paid ten thousand Yemeni rials (equivalent to 50 US dollars) by the university. The test is administered one month before the beginning of the academic year in order to give the unselected applicants a chance to join other departments of their choice.

The number of students intending to join the faculty of Education every year ranges between 500 and 600 students. Each 50 to 60 students take the test in a separate examination hall and are observed by 4 to 5 proctors. The number of students intending to join the Faculty of Arts, on the other hand, is always less and ranges between 300 and 500 students, distributed between ten examination halls. Each group of students is also observed by 4 to 5 proctors. The proctors are non-teaching staff members of the university and they usually have little knowledge of English.

Each admission test runs into more than three pages and the answers are usually written on the question papers. There used to be an oral test to supplement the written test but this practice has been done away with since 1995 due to time constraints and the increasing number of applicants. Since then the applicants have had to pass only the written test, which has often been a three-hour test. The answer sheets are then scored by a committee of 5 to 6 teaching staff, holders of at least a bachelor degree in English, headed by the chairperson. The scoring of the test usually takes three to four days and the results are announced two to three days later. The number of students selected depends on the incoming capacity of the departments concerned. For example, if the English department at the Faculty of Arts requires a hundred students, the committee would select the top hundred students, regardless of their scores.

**Statement of the Problem**

English admission tests at Taiz University are constructed by any teaching staff holding a PhD in English or linguistics. The results of these tests are used to make selection decisions which have a life-long impact on the lives of the individual test takers and a long-term impact on the programs of instruction. Assigning the task
of test construction to teaching staff with little professional expertise and knowledge of testing creates, or has created, many problems at the levels of test design, test operationalization, and test administration and use.

On examining ten admission tests administered between the years 1995 and 2005, the researcher noticed a number of test internal inconsistencies. Many of the tests, for example, did not include questions on writing composition. The listening skill, though it is very important at later stages of university education, was not tested at all. The test tasks were partly indirect. They tested reading and writing abilities directly but oral interaction was tested indirectly through written dialogues and at times it was not tested at all. The test tasks were also not fully authentic. They did not fully reflect communicative activities which the candidates were likely to engage either in personal or in academic situations.

The marks distribution was also problematic. Many of the tests allocated twenty marks for the reading, writing and translation questions each. The remaining forty marks were distributed among four types of questions: plural formation, choice of prepositions, tense, and negation. All these questions test knowledge of grammar. In other words, the tests were grammar-based proficiency tests. The ability to manipulate the grammar of English and the ability to manage reading comprehension passages were considered the most important abilities and together given more than half the total mark.

Grammar-based proficiency tests are limited in terms of the information they can provide about a candidate. The results of these tests cannot lead to a generalization that the candidates who passed the tests are good at English. Good performers in these tests may well be poor communicators. Only communicative tests can provide all the information necessary to determine whether or not a candidate has the necessary language abilities required to cope with the future course demands.

The primary purpose of this study is to initiate a professional admission testing program at Taiz University. Admission test tasks should reflect the target language use domain and should, as far as possible, test all language abilities in communicative situations. The test writer should also specify the purpose of the test, the characteristics of the test takers, the characteristics of the target language use domain (or TLU domain), etc. The students’ test scores on test sections and test items should be tabulated and stored. All this information is necessary in the evaluation of tests and for gradually creating more reliable, valid, and effective admission tests. The first part
of this study makes a quantitative analysis of the 2005 admission test using Classical Item Analysis and other statistical means and a qualitative analysis of the admission testing situation at Taiz University. The second part of the study constructs an admission test (in its three stages of design, operationalization, and administration and use) based on the model of communicative competence proposed by Bachman and Palmer (1996).

**Significance of the study**

Admission tests are one of the most high-stakes tests. Great care should be taken in the construction and administration of these tests and in the use of their results. At Taiz University, admission tests have, ever since the establishment of the university, been produced without analyzing their results and without benefiting from previous test administrations. An unfortunate consequence was that data was wasted and opportunities to gradually create more reliable, valid, and effective tests were also lost.

The present study in its two stages of evaluation and construction presents an attempt to create a set of practical procedures which could be used in the production and evaluation of in-house English admission tests at Taiz University. The first part of the study, the evaluation part, uses data from a previous test administration to investigate the reliability and effectiveness of that test. The researcher uses the SPSS and Microsoft Excel software to calculate the descriptive statistics and item statistics of the test and of the test items respectively. This quantitative analysis, or statistical analysis, of test data is carried out for the first time in the context of admission testing at Taiz University. The analysis, however, is restricted to the descriptive statistics, to the reliability of the test and to the item statistics of the test items. The validity of the test is not investigated because information necessary for conducting a validation study is not available. The second part of the study, on the other hand, attempts to construct a test which has all the necessary information which will be required in any subsequent evaluation study of the test, especially a validation study. The test is based on a model of communicative competence proposed by Bachman and Palmer (1996).

The study is significant because it constitutes a first attempt to evaluate test data at Taiz University and a first attempt to introduce a theory-informed admission test. These are crucially necessary steps towards the development of reliable and effective selection instruments of testing language proficiency. The evaluation of tests helps test developers identify the weaknesses and strengths of tests and improve
subsequent tests. By ensuring that only the candidates with adequate language proficiency are admitted into the courses, the students, teachers and the whole education system all stand to gain.

This study may not add to the theory of language testing but it extends the existing literature on proficiency test evaluation and construction. It also provides a much needed, and long overdue, framework for the construction and evaluation of subsequent admission tests at the English departments of Taiz University. The study is also among the first to examine the applicability of the Bachman and Palmer (1996) model in the context of admission testing in Yemen. It is a first step towards further research in the area. The findings of this study, however, are not confined to Taiz University. Since communicative language testing has world-wide applicability, the findings can be used in the improvement of similar admission test situations in similar EFL contexts.

Organization of the Dissertation

This dissertation is organized as follows. Chapter One is an introduction. It starts with stating the position of language testing in Applied Linguistics and how it has emerged as an independent discipline within Applied Linguistics. This is followed by a brief discussion of the development of language testing during the last three decades. The context of the present study is also introduced. The chapter includes a brief overview of Yemen the country, the status of English Education in Yemeni schools and universities, and of the status of English educational assessment in Yemeni schools and universities. The overview ends with a cursory look at the English admission test situation at Taiz University. This is followed by an outline of the problem to be investigated, the significance of the study, a summary of the five chapters of the dissertation and the definition of key terms and concepts in language testing which are relevant for the work.

Chapter Two is a review of the related literature on language testing. The review is in two parts: a) approaches to language testing, and b) methods of statistical analysis. It begins with a discussion of the main approaches to language testing, namely, discrete point testing, integrative testing and communicative language testing, with special focus on the principles of communicative language testing and how they differ from those of the other approaches. The second part of the review discusses the
methods of statistical analysis that are used in the quantitative study of the 2005 admission test.

Chapter Three provides an analysis of the 2005 in-house admission test using two types of analysis: quantitative analysis and qualitative analysis. The quantitative analysis includes three types of statistical analysis: descriptive statistics, classical item analysis (CIA), and the ‘Cronbach’s Alpha if Item Deleted’ analysis. The qualitative study includes analyzing three questionnaires given to the test takers, the test writer, and the test scorers and raters. This analysis is conducted to diagnose the weaknesses of the test itself in particular and of the testing situation in general.

Chapter Four is an attempt to develop a test which incorporates features of the discrete point testing approach, integrative testing approach and communicative language testing approach. The test is constructed with a view to meeting Bachman and Palmer’s six qualities for test usefulness. It is constructed according to their organization of test into three stages: design, operationalization and administration.

Chapter Five is the concluding chapter. This chapter sums up the findings arrived at in the previous chapters and discusses these findings with regard to the admission testing situation at Taiz University. It also makes recommendations for improving the three stages of admission test development, namely, test design, operationalization and administration, giving due attention to the techniques of scoring and the assessment of the test usefulness for its intended purpose. This chapter ends with the limitations of the study and its suggestions for further research.

**Definition of Terms**

**Qualities of Test Usefulness**

**Reliability**

Reliability refers to the consistency of measurement, whether the test is administered in different situations or scored by different scorers.

**Construct Validity**

Construct validity pertains to the meaningfulness and appropriateness of the interpretations that we make on the basis of test scores, and involves providing evidence justifying the interpretations.
Authenticity
Authenticity is the degree of the correspondence of the characteristics of a given language test task to the characteristics of a TLU (Target Language Use Domain) task.

Interactiveness
Interactiveness is an important test quality because it pertains to the degree to which the constructs intended to be assessed (i.e., language knowledge plus strategic competence) are critically involved in accomplishing the test task.

Impact
Impact can be defined broadly in terms of the various ways in which test use affects society, the education system, and the individuals within these (for example, test takers and test users).

Practicality
A practical test is one whose design, development, and use do not require more resources than are available – human resources, material resources, and time.

Cloze Test
A test in which, after an introductory sentence which is left intact, words in a reading passage are systematically removed – every 5th, 6th, or 7th word was a typical procedure – and replaced with a blank.

TLU Domain (Target Language Use Domain or The Domain of Generalization)
TLU domain is a set of specific language use tasks that the test taker is likely to encounter outside of the test itself, and to which we want our inferences about language ability to extrapolate (Bachman and Palmer, 1996, p.44). The tasks in the TLU domain are referred to as ‘TLU tasks’.

High-stakes Tests
High-stakes tests are major, life-affecting, tests which provide information on the basis of which significant decisions about candidates are made, e.g. admission to courses of study or to work settings.

Norm-referenced Test (NRT)
A norm-referenced test is a type of test “whereby a candidate’s scores are interpreted with reference to the performance of the other candidates. Thus the quality of each performance is judged not in its own right, or with
reference to some external **criterion**, but according to the standard of the group as a whole” (Davies, et al., p. 130 [original emboldening]).

*Criterion-referenced Test (CRT)*

A criterion-referenced test is a test that “examines the level of **knowledge** of, or **performance** on, a specific **domain** of target behaviours (ie the **criterion**) which the candidate is required to have mastered”. Criterion-referenced test scores “report a candidate’s **ability** in relation to the criterion, i.e. what the candidate can and cannot do, rather than comparing his/her performance with that of other candidates in the relevant population, such as happens in **norm-referenced tests**” (Davies et al. 1999, p. 38 [original emboldening]).