CHAPTER I

INTRODUCTION

Rationale and Significance of the Problem

Assessment is one of the most important parts of language learning and teaching. It can helps learners to know where they are, how much they have achieved and how much they still need to build up their language skills. It can help to find out how much learners have learnt after a certain time of period. Assessing learners' level and ability can be done in many different ways and formats. Types of assessment can be varied according to the purposes of the language program and the assessment.

In many language programs, tests are likely to be used for assessing learners' achievement as tests can show the progress of learners as well as help learners to improve their skills. Madsen (1983, p.4) claimed, "Good English tests help students learn the language by requiring them to study hard, emphasizing course objectives, and showing them where they need to improve". The content and kind of test can be varied according to the purposes of the tests. However, there are some common important qualities that every test should have. Tests should be valid, reliable and not too difficult or too easy for the candidates. Testing in language programs should be carefully designed to find out what the students have learned and how much learners can use the target language. Validity and reliability are important for every test.

In most language programs, achievement tests are designed to assess learners' progress over a certain amount of time. There are many factors which need to be

considered in writing achievement tests. Tests should have specific objectives in testing learners. As tests are used to measure the students' achievement over the objectives of the program, tests used in language program should have the components of a good test, like validity and reliability. Without these two qualities, there is no point of testing.

However, not all achievement tests used in language programs possess these qualities. Some language programs still have weaknesses in developing good tests for learners because of factors such as lack of knowledge, limited experience, resources and financial support. In many language programs, the teachers' lack of awareness of important components of language testing like reliability and validity create invalid and low quality tests for the learners and the program. Even if the program and the test developers are aware of the test qualities, they may not be able to put this awareness into practice for reasons such as limited resources and financial support. There may be many other reasons that hinder programs from developing a good test.

Therefore, achievement tests used in language programs should be evaluated. Richards (2001, p.288) claimed that "Evaluation may be carried out as part of the process of program development in order to find out what is working well, and what is not, and what problems need to be addressed." Evaluation can give information on what aspects of the program are good for language learning and teaching, and what are not. From that, teachers, curriculum developers, and school authorities can consider improvement of the program. Genesee (2001, p.144) claimed the importance of evaluation in leading to ".. informed decisions that enhance student achievement and the success of educational programs". This study is an evaluation of an achievement test with the aim of improving this aspect of the language program.

2

Background to the Study

Achievement tests are widely used to assess learners' language ability in Myanmar. However, language teacher education in Myanmar is limited and not accessible for all teachers. Many teachers teach English language without proper training about teaching. Because of limited knowledge of the field of language teaching and learning, some teachers are not aware of important factors like validity and reliability in language testing. Tests are sometimes developed without objectives or test specifications. Important key factors in testing like validity and reliability of tests are not considered by the test developers in some situations. This leads to poor quality tests. Therefore, this provided a personal motivation to evaluate the achievement test used in one language program of Myanmar in order to find out what strengths and weaknesses exist in the test construction.

The achievement test of the Bachelor of Arts in Religious Studies (BARS) program at Myanmar Institute of Theology, Yangon, Myanmar is evaluated in this study. The BARS program was established in response to the challenge of the education situation in Myanmar. Its main purpose is to provide a Bachelor-level education to students, regardless of ethnicity, sex, age, marital status, religion and physical disability. The program offers a broad-based liberal arts curriculum that includes religion, philosophy, the humanities, social science, business, mathematics, computer science and English language.

The BARS program is a four-year bachelor degree program and offers six different majors- English, Computer Science, Business Management, Theology, Social Study and

3

Fine Arts. In this program, students have to take English as a compulsory subject from the first year to the third year. School authorities choose what to teach throughout three years. The English course has been using *New Headway* English course book series starting from pre intermediate level to advanced level for four skills course and an in-house vocabulary series.

Every term, the students have to take an achievement test for the English course at the end of the semester. The assessment system is designed by the school authorities and some teachers. This program uses tests which are created by a number of the English teachers. Test content is mainly based on the materials found in the syllabus. In this program, no systematic evaluation of the test has been carried out. The test construction process, and evaluating the validity and reliability of the test content are important parts of developing a good test. These test factors need to be evaluated in order to establish the quality of the test and to help the program to improve to create a better test for the learners for the improvement of the program.

Objectives of the Research

The broad aim is to evaluate the achievement test used for an English course in this bachelor degree program in the areas of the test construction process, the content validity of the test and the reliability of the test.

The questions that are explored are as follows:

1. What kind of processes is involved in construction of the test?

2. How valid is the content of this test for the purpose of assessing achievement?

3. How reliable is the achievement test being used in this program?

The research questions are related to one another as the second and third question address different components included in the first question.

From the research questions, if the test developers do not have specific objectives for their achievement test, the test may not be able to achieve its aim and meet objectives of the program. This can be evidence suggesting that test developers need more knowledge about language testing. However, if the test developers provide objectives, specifications of the test, criteria level of performance and clear scoring procedures, it may be possible to say that the test is well constructed and it can be cited as a good test which can really assess the learners' ability.

Significance of the Study

The findings obtained from this study will be of interest to a number of groups. This study will provide information about the validity and the reliability of the achievement test currently used in the bachelor program offered at the Myanmar Institute of Theology. Therefore, the first group interested in the study is the administrators of the program itself. This study will help them to see the quality of the achievement tests used currently and it will provide an insight into the validity and reliability of the tests. From that, the program authorities, test developers and teachers can work together in order to produce

higher quality tests for the development of the program. Moreover, the study will help the

program authorities evaluate the need for teacher education.

The next group is the test developers or teachers who design the test. The

findings will provide them with information about the quality and the validity of the test

and what they should change in order to improve the test.

Moreover, the curriculum developers for this program, other curriculum

developers and test developers, and other English language institutions in Myanmar and

beyond could benefit from this study, as this study will present an example of an

evaluation of test construction of an in-house achievement test.

Definition of terms

Achievement tests   Achievement tests are used for assessing

                                     learners' ability over a period to

                                     measure how much learners have learnt

                                     from what has been taught. The main

                                     purpose of an achievement test is to find out

                                     whether or not course objectives have been

                                     achieved Alderson (1982, p.217) claimed, "The

                                     results of the achievement tests can be used

                                     by teachers to slow down or speed up the

                                     pace of instruction, to change the textbook,

                                     to supplement materials with different

activities and so on." There are two kinds of achievement tests: final achievement tests and progress achievement tests.

Essay writing

Essay writing is a type of test item where a test taker is required to write an extended piece of a text on a set topic (Richards, 2002).

Evaluation

Brown (1995, p.218) defines evaluation as "the systematic collection and analysis of all relevant information necessary to promote the effectiveness of a curriculum and assess its effectiveness within the context of the particular institution involved". It is looking through what has been done or looking at what is working well or not and knowing how it works. There are three different purposes of evaluation namely normative, illuminative and summative. Formative evaluation focuses on on-going development and improvement of the program and it may be carried out as part of the process of the program. Illuminative evaluation seeks to

find out how different aspects of the program work or are being implemented. Summative evaluation investigates the worth and value of different aspects of the curriculum and the effectiveness of the program.

Gap- fill   Items in which the candidates have to fill a gap with a word are called 'gap-fill' items. Gap- fill items are used in assessing reading, listening, grammar and vocabulary (Hughes, 2003).

Matching   Matching is a type of test item that requires the candidates to indicate which entries (e.g. words or phrases) on one list is the correct matches for the entries on another list (Richards, 2002).

Multiple choice   Multiple choice is a test item format where the candidate is confronted with a question together with three or more possible answers from which the correct answer must be selected. Usually, the first part of a multiple choice item will be a question or an incomplete statement, and this is known as a stem. The different possible answers provided are known as alternatives. The alternatives contain

	(usually) one correct answer and several wrong
	answers or distractors (Richards, 2002).
Process of test construction	The process of constructing a test is one
	of the fundamental component in test
	development. Constructing a test is much
	more than writing test items. It involves
	setting objectives of the test, drawing test
	specifications, writing test items, informal
	trialing of test items, validating and
	moderating test items before administration.
Reliability	Reliability of a test is defined by Richards (2002,
	p.454) as "a measure of the degree to which a test
	gives consistent results. A test is said to be reliable
	if it gives the same results when it is given on
	different occasions or when it is used by different
	people."
Short answer questions	The Open-ended items, where the candidates have
to	think up the answer for themselves are called 'short
	answer questions'. The answers may range from a
	word or phrase to one or two sentences. This is a

common technique for assessing reading and
listening (Alderson, Clapham & Well, 1995)

True/false

True/false items are a type of items or test task that
requires the test takers to decide whether a given
statement is either "true" or "false". True/ false
items are sometimes modified by requiring
candidates to give a reason for their choice. True/
false items are more suitable for classroom progress
and achievement tests. (Richards, 2002; Heaton,
1988)

Validity

Validity is one of the components of a good
test. Richards (2002, p.575) defined validity as "
the degree to which a test measures what it
is supposed to measure, or can be used
successfully for the purposes for which it is
intended". Every test needs to be valid. Reliability
and validity are interrelated and both are essential
components of a good test.

Organization of the Study

Chapter One has introduced the study, and included a rationale and discussion of the significance of the problem. This chapter has also provided the objectives of the research and definition of terms. Chapter two will review some of the background literature and discuss the process of test construction of the achievement test and important components of tests including validity and reliability. Chapter three presents the research methodology. After that, data analysis is presented in chapter four. Finally, the discussion and conclusion of the study are given in the last chapter. Suggestions for constructing good quality tests and suggestions for further research are also presented in this last chapter.