۸۱/۱۰/۱۴

T.M.U

Tarbiat Modarres University
Faculty of Humanities
English Department

On the Reliability of T.M.U English Examination

A thesis submitted in partial satisfaction of the requirements for the degree of master of arts in the teaching of English as a Foreign Language(TEFL)

By
Malakeh Haghighi

Supervisor:**Dr. Gh. Kiany**

Advisor:**Dr.Akbar Mirhassani**

March, 2002

۴۱۹۴۹

IN THE NAME OF GOD

٤١٩٤٩

We Recommend This Thesis by Malakeh Haghighi Entitled

On the Reliability of T.M.U English Examination

Be Accepted as Partial Fulfillment of the Requirement for the Degree of Master of Art in Teaching English as a Foreign Language (TEFL).

Committee of Final Examination:

..........................Dr. Gh. Kiany	Supervisor, and the Head of the Department
..........................Dr. A. Mirhassani	Advisor
Mohammad Alavi ....Dr. M. Alavi	Reader
..........................Dr. R. Ghafarsamar	Reader

Tarbiat Modarres University

Tehran, Iran

March, 2002

Dedication:

To My Kind Family

ACKNOWLEDGEMENTS

I would like to express my heartful thanks and appreciation to Dr.Kiany, the supervisor of this thesis, who patiently guided me throughout the preperation and completion of this study. Obviously, without his precious comments and suggestions the whole body of this study would not have been accomplished. I should also thank Dr. Mirhassani, the advisor of this study, for his tolerance in reading the thesis and giving helpful comments.

My further thanks to Dr. Alavi and Dr.Ghaffarsamar who carefully read this thesis.

I am also deeply indepted to Dr. Salimizadeh from Allameh Tabatabaee University, faculty of Educational Sciences who generously provided the computer programs for IRT analyses and gave me a great deal of help in running the utilized programs.

I am also thankful to my friends and colleagues, especially Miss. Nazarpur and Mr. Salimi, who have encouraged and assisted me during the fulfillment of this work.

This thesis doubtlessly has still many faults, and I must acknowledge them as my own.

ABSTRACT

The majority of the Iranian universities, which offer doctoral programs, have started to accept only those candidates who have a good command of English proficiency which is usually measured either by a general English test administered by the same university or by tests like MCHE, TOLIMO, MOHMET which are administered by the related ministries. Results of the these tests are used to make very important decisions on the fate of candidates; hence these test should be carefully investigated in terms of reliability, validity, and the other characteristics of the test. T.M.U English Examination is a TOEFL-like English language proficiency test used as a prerequisite for Ph.D Entrance Examination at Tarbiat Modarres University. In Esfand,1379, 6000 male and female postgraduates from different universities and different fields took part in this exam. This study attempted to investigate reliability, test difficulty, and speededness. To achieve the purpose of the study, the data obtained from about 2000 male and female Ph.D applicants were analyzed through different types of reliability estimates. The reliability coefficients of the total test and its subparts were estimated through KR-21, cronbach alpha, and split-half reliability. Then the reliability coefficients of different types of reliability were compared for significant differences. Further, the test difficulty as a general characteristic of the test, which may influence the size of reliability estimates, was also investigated. Related to the issue, the difficulty and discrimination values of the test items were calculated through classical item analysis and IRT methods. This study also investigated speededness as a source of measurement error. Two techniques were used in an attempt to investigate speededness. One of these techniques provided an estimate of the degree to which all examinees truly reached the 75 percent on the test, and the second determined if 80 percent of examinees truly completed the last set of items. The results revealed that though T.M.U English Examination was to some extent reliable, the reliability coefficients did not meet the levels which would be required for a high-stake test. Also it was found that based on norm-referencing criterion T.M.U exam did not have acceptable level of difficulty.That is, it was difficult for the examinees. Further the results showed that the test was slightly speeded as far as the ETS criteria are considered. Results of this study may be of benefit to test developers in general and language test designers in particular. More specifically, policy makers and test developers of T.M.U, Ministry of Science, Research and Technology, Ministry of Health and Medical Education, and other such institutes may get some benefit from this study.

TABLE OF CONTENTS

Title Page

CHAPTER III : METHOD 66-71

CHAPTER IV : RESULTS AND DISCUSSION 73-101

LIST OF TABLAES

CHAPTER I

INTRODUCTION

CHAPTER 1
INTRODUCTION

1.1. OVERVIEW

A language proficiency assessment attempts to measure a person's ability to understand and produce the language. It may be a global assessment covering the four major skills, i.e. listening, reading, writing, and speaking; or it may focus on only one or two of these skills. A proficiency test which is used to place students in a language program, to exit them from the program, or to admit them to other educational or employment opportunities should possess the basic characteristics of a test. If we are to interpret the score on a given test as an indicator of an individual's ability, that score must be both reliable and valid. These qualities are thus essential to the interpretation and use of measures of language abilities, and they are the primary qualities to be considered in developing and using tests. While validity is the most important quality of test use, reliability is a necessary condition for validity, in the sense that test scores that are not reliable can not provide a basis for valid interpretation and use (Bachman, 1990).

According to Walsh and Betz (2001), "the first requirement for a high-quality, or 'good' test is that the test possess what is called reliability" (P.47). Reliability is the extent to which test scores are free from errors and hence consistent (Alderson et al. 1995). In examining reliability we must identify potential sources of measurement error and estimate the magnitude of their effects on test scores. Errors of measurement, or unreliability, should be concerned, because test performance is affected by factors other than the abilities to be measured (Bachman, 1990). According to Black (1998), in most assessment systems, serious attempts are made to minimize the possible effects of these factors, but some are hard to deal with and it is inevitable that there will be a degree of error in any result. For instance, there are factors such as poor health, fatigue, lack of interest or motivation, and test–wiseness that can affect individuals' test performance, but which are not generally associated with language ability.

In addition to unsystematic factors, according to Bachman (1990), the test method facets are also potential sources of errors that can affect the accuracy of the device in measuring language abilities. Among the test method facets, there is test rubric, which consists of the facets that specify how test takers are expected to proceed in taking the test. The amount of time allocated for the test or its parts is likely to affect test performance, too. In some tests, the time limit is such that not all test takers can manage to answer all the items or parts of the test. In these tests, which are speeded tests, test scores are partly a function of the test taker's level of ability and partly a function of the speed or rate at which individuals can complete the test.

In this regard, testing theorists distinguish between tests that measure power and tests that measure speed. Items in a pure power tests range in difficulty, and there is no time limit: the goal is to measure how accurately examinees can answer the items. Items in a pure speed test are very easy, and the time limit is strict: the goal is to measure how quickly examinees can respond (Schnipke & Scrams, 1997).

Accordingly, speededness in testing is the effect of time limits on the test–taker's scores. An exam is speeded to the extent that those taking it score lower than they would have if they had unlimited time.

Speededness is often measured by calculating the proportion of examinees who do not reach a certain percentage of test items (Ibid).

Speededness is a problem for test theory. Gulliksen (1950) points out that the item indexes of classical test theory were developed for power tests and might not be appropriate if the speed component is too large. Similarly, Hambleton and Swaminathan (1985) argued that unidimensional item response theory (IRT) models implicitly assume that the test is unspeeded: speed and power components would require separate dimensions. Additionally, Oshima (1994) and Schnipke (1996) showed empirically that IRT item parameter estimates are distorted by speededness.

Besides investigating measurement errors, the analysis of individual items is of fundamental concern in the development and use of language tests. Item analysis includes principally the measurement of item difficulty and item discrimination. Both the validity and the reliability of any test depend ultimately on the characteristics of its items. High reliability and validity can be built in to a test in advance through item analysis (Anastasi, 1982).

This study represents an attempt to determine if the English Examination of Tarbiat Modarres university is a reliable test for its