In the Name of God

١٥٠٨٥٥

University of Isfahan
Faculty of Foreign Languages
Department of English

M.A Thesis

Investigating the Construct Validity of the FCE Reading Paper in the Iranian EFL Context

Supervisor
Dr. Hossein Barati

Advisor
Dr. Hossein Pirnajmuddin

By
Elahe Tavakoli

۱۳۸۷/۱۰/۲۸

October 2007

۱۰۰۸۰۵

دانشگاه اصفهان

دانشکده زبانهای خارجی

گروه زبان انگلیسی

پایان نامه ی کارشناسی ارشد رشته ی

زبان انگلیسی گرایش آموزش

بررسی سازه روایی آزمون(FCE)– بخش خواندن و درک مطلب– برای فراگیران زبان انگلیسی درایران

استاد راهنما:

دکتر حسین براتی

استاد مشاور:

دکتر حسین پیرنجم الدین

پژوهشگر:

الهه توکلی

۱۳۸۷ /۵/ ۲۸

مهر ماه۱۳۸۶

۱۰۵۸۵

To my parents and the ones who supported me throughout this research

Acknowledgment

At the completion of this study, I feel indebted to a number of people whose support was always available to me.

In particular, my sincere appreciation and thanks go to Dr. Hossein Barati, my supervisor, who was an outstanding example of supervision and support. His guidance was practical, rigorous and insightful. I was very fortunate and privileged to have his counsel and guidance available to me. Without his invaluable sources of research, critical questions and insightful comments, the completion of this thesis would be impossible.

I also owe a great debt of gratitude and appreciation to Dr. Pirnajm al-din, my advisor, whose help, guidance, advice and encouragement were always invaluable to me during this study.

Further, my deep gratitude should go to Dr. Michael Milanovich, the Chief Executive of the UCLES validation Committee and Dr. Ardeshir Geranpayeh, UCLES' senior coordinator, for their consultancy on the FCE and its administrations in Iran.

I am also grateful to all of the lecturers at Isfahan University and Azad University of Najafabad who so kindly contributed to this study as expert judges and those who permitted me to use their classrooms for data collection.

I also feel grateful to all the undergraduate students who participated in various sessions of data collection for this study.

Last but not least, my heartfelt gratitude and special consideration should go to my dear parents for providing me with unconditional support and encouragement.

Abstract

Validation studies on language proficiency tests have attracted many researchers in the last decades. Such studies investigate the skills under assessement by the items in the tests. The present study aims to investigate the construct validity of the reading paper of the First Certificate in English (FCE) in the Iranian EFL context. This research addresses the following three questions: (1) Do the majority of the Iranian EFL expert judges agree on the skills measured by the items in the FCE reading paper? (2) Do the majority of the Iranian EFL undergraduate test-takers come to agreement on the skills measured by the items in the FCE reading paper? And (3) Do items in the FCE reading paper assess the same reading skills claimed by the test's developing board, i.e. UCLES, in the context of Iranian EFL undergraduates?

The study deploys a triangulated approach in collecting and analysing the data. It attempts to combine both qualitative (i.e. experts' and test-takers' judgmental approach) and quantitative (i.e. Factor Analysis) data analytic methods to address the above questions

The findings at the judgmental phase revealed that there was not a considerable degree of agreement among the judges on the skills assessed by the total FCE reading paper or its items. Considering the EFL test-takers, no significant agreement was observed amongst the test-takers on the skills measured by the total FCE reading paper or its items.

The study finally concluded that the quantitative findings of the factor analysis revealed similar outcomes with the judgmental phase. They indicated that the individual items in the FCE reading paper do not assess the set of reading skills claimed by the test developers. In other words, the reading which is assessed by the FCE reading paper has a unitary nature rather that componential.

Key Words: Construct, Reading Construct, Unitary Nature of Reading

Table of contents

Title **page**

CHAPTER1: INTRODUCTION

1.1. Overview.....1

1.2. Statement of the Problem 2

1.3. Significance of the Study.....2

1.4. The Purpose of the Study 4

1.5. Research Questions and Hypotheses.....5

1.6. Definition of Key Terms.....6

1.7. Outline of the Study 7

1.8. Summary.....8

CHAPTER 2: REVIEW OF LITERATURE

2.1. Overview.....10

2.2. EFL Proficiency Tests.....11

2.2.1. Test of English as a Foreign Language (TEOFL).....12

2.2.2. International English Language Testing System (IELTS).....14

2.2.3. Cambridge Main Suit Examinations.....16

2.3. The First Certificate in English (FCE).....17

2.3.1. The FCE before 1939.....17

2.3.2. The FCE after 1939.....18

2.3.3. The FCE Format 19

2.3.4. The FCE scoring.....21

2.4. Review of the Related Research on the FCE 22

2.4.1. The FCE and the TOEFL.....23

2.4.2. The FCE and the IELTS.....26

2.4.1. How Does UCLES target the Reliability of its examinations?.....29

2.4.2. How Does UCLES targets the Validity of its examinations?.....30

2.5. Validity 31

2.5.1. Priori Validity.....32

2.5.1.1. Theory-based Validity.....32

Title **Page**
2.5.1.2. Context/ Content Validity....33
2.5.2. Posteriori Validity....33
2.5.2.1. Scoring Validity....33
2.5.2.2. Criterion/ Empirical Validity....33
2.5.2.3. Consequential Validity....34
2.5.3. Face Validity....34
2.6. Construct Validity....35
2.7. Reading Construct....36
2.8. Testing Reading Skills....37
2.8.1. Unitary Construct of the Reading Comprehension....39
2.8.2. Componential/ Divisible construct of Reading Comprehension....40
2.8.3. Bi-divisibile Construct of Reading Comprehension....43
2.9. Studies on the Reading Skills....44
2.9.1. Alderson and Lukmani (1989)....44
2.9.2. Alderson (1990a)....45
2.9.3. Alderson (1990b)....46
2.9.4. Farhady and Hessamy (2005)....47
2.9.5. Lumley (1993)....48
2.10. Summary....50

CHAPTER 3: METHODOLOGY

3.1. Overview....51
3.2. Research Questions and Hypotheses....51
3.3. Research Design....52
3.4. The Pilot Study....53
3.4.1. Aims....53
3.4.2. Participants....54
3.4.3. Procedure....54
3.4.4. Implications for Main Study....55
3.5. The Main Study....57
3.5.1. Methodology....58

Title **Page**

3.5.2. Participants 58

3.5.3. Instrumentation 59

3.5.3.1. The FCE Reading Paper.....59

3.5.3.2. The EFL Taxonomy of Reading Skills 60

3.5.4. Procedure 61

3.6. Summary 63

CHAPTER 4: DATA ANALYSIS AND RESULTS

4.1. Overview.....64

4.2. Normality.....65

4.3. The Reliability of the FCE.....66

4.4. Addressing Research Hypothesis 1.....67

4.5. Addressing the Research Hypothesis 2.....70

4.6. Addressing the Research Hypothesis 3.....72

4.7. Summary 74

CHAPTER 5: DISCUSSIONS, IMPLICATIONS AND CONCLUSIONS

5.1. Overview.....75

5.2. Restatement of the Problem 76

5.3. Discussion of Findings.....77

5.3.3. RH1: EFL Expert Judges & the FCE Skills Identifiability.....77

5.3.2. RH2: EFL Undergraduates & the FCE Skills Identifiability 79

5.3.3. RH3: Quantitative Factor Analysis & the FCE Skills Identifiability.....81

5.4. Conclusion.....84

5.5. Limitations of the Study.....85

5.6. Implications of the Study.....87

5.7. Suggestions for Further Research.....88

Title **Page**

List of Appendix

Appendix 3.1: Normal Distribution of Variables: (FCE) Pilot Study.....91

Appendix 3.2: The EFL taxonomy of reading skills (Barati2005).....92

Appendix 3.3: The FCE test.....93

Appendix 4.1: Normal Distribution of Variables: (FCE) Main Study.....103

Appendix 4.2: Test of Factorability of data, FCE: KMO and Bartlett's Test 104

Appendix 4.3: Test of factorability of data: (FCE) 105

Appendix 4.4: Screeplot.....106

List of References.....107

List of Tables

Title **Page**

Table 2.1: Cambridge Main Suit and ALTE Levels.....16

Table 2-2: The FCE Reading Paper: focus and test methods.....19

Table 2.3: Reliability estimates for the FCE reading paper 2000-3 29

Table 2.4: VIRP- based checklist
(section on validity, extracted from Saville 2002: 9)..... 30

Table 2.5: Weir's (1993) Summary Checklist of Operations in Reading 38

Table 2.6: Weir's (1997) Revised Summary Checklist
of Operations in Reading 40

Table 3.1: The distribution of the participants'
age, gender and level of education in the quantitative study..... 57

Table 4.1: Distribution of variables, FCE total 65

Table 4.2: Reliability Estimates, The FCE reading paper:
its total and sub-tests and total 66

Table 4.3: Expert Judges' inter rater reliability.....67

Table 4.4: frequency of the EFL experts' decisions
on the item/ skill correspondence 68

Table 4.5: frequency of the EFL experts' decisions on
the item/ skill correspondence.....70

Table 4.6: Component Matrix for the FCE reading paper.....72

List of Figures

Title **Page**

Figure 3.1: Score distribution and normal curve, for FCE reading paper.....55

List of Acronyms

EFL: English as Foreign Language

FA: Factor Analysis

FCE: First Certificate in English

UCLES: University of Cambridge's Local Examination Syndicate

Chapter One: Introduction

1.1. Overview

The concept of test validity has attracted the attention of many language test constructors in the last few decades. The tendency towards validation studies has been more evident in relation to universal language tests of proficiency. The present investigation explores the construct of reading, with a specific reference to the reading skills which are claimed to be assessed by the FCE reading paper. The study considers the use of the FCE reading paper in the context of Iranian adult learners of English as a foreign language (EFL). In other words, the aim, here, is to investigate the reading skills deployed by Iranian EFL undergraduates while completing items in the FCE reading paper and to compare the aforementioned skills with those claimed by the FCE constructors (UCLES). The study adopts both qualitative and quantitative approaches to validate the FCE reading paper in terms of the skills claimed to be assessed by its items.

The first chapter; therefore, is an introduction to the present research on the FCE reading comprehension skills and begins with the existing problems in testing reading skills (1.2) and the significance of the study (1.3). The purpose of the study will be discussed in 1.4. In the consecutive sections, research questions and the hypotheses (1.5) together with the definitions of some key terms (1.6) are

presented. The last section of this chapter provides an overview of the structure of this thesis (1.7).

1.2. Statement of the Problem

So far test constructors have devised different tests based on different purposes; however, they have to make sure about the abilities being measured by individual items in their tests. This corresponds to the validity of the test (i.e. what the test purports to measure). In terms of reading comprehension tests, validity has been a matter of great concern and several researchers have posed related issues and problems when talking about the abilities/skills measured by the individual items in reading tests (e.g. Alderson 1990a, 1990b).This is mainly due to the nature of reading: whether it is unitary (e.g. Alderson and Lukmani 1989, Alderson 1990), or componential (e.g. Lumley 1993, Weir and Porter 1994) (see 2.8).

When assessing language reading comprehension ability, it seems vital to define first what exactly reading comprehension ability entails. However providing an adequate definition of reading comprehension ability is not an easy task. The task seems more difficult when the items in a test of reading comprehension are assessed for their construct validity; moreover, the controversial views on the unitary or componential nature of reading comprehension contribute to this difficulty. The idea of unitary or componentiality of reading concerns the skills comprising the reading abilities. It is suggested that for the benefit of teaching and testing, a unitary view of reading should be discarded since the process of reading involves the use of different skills and strategies (e.g. Weir *et al.*, 2000). A number of language learning/reading practitioners (e.g. Munby 1978) proposed taxonomies of reading skills believing that those skills comprise the ability of reading; however, in language testing there

are few reading-skills taxonomies which have appeared influential (e.g. Munby 1978, Lumley 1993). Many of the other taxonomies have not been empirically investigated for their validity and are viewed as mere armchair speculations (e.g. Munby 1978).

It has often been argued that the key issue is not how many reading skills we can think of, but how many can be shown to be measured by the items in a test. However, there has been considerable disagreement about the number of skills assessed by a test of comprehension, and the statistical techniques which are used to analyze the data (Alderson, 2000). Other factors such as first language, background knowledge, system of education and interest have made the assessment of reading skills more complicated. Thus, it seems important to seek validity evidence for the reading comprehension tests i.e. which skills are measured by any test of reading comprehension.

Probably the problem becomes more evident when it comes to administering universal tests such as TOEFL, IELTS, EAP, ESP and the FCE in a foreign language context. Although these tests claim to be standardized and widely used in many countries around the world, there has not yet emerged any clear-cut evidence as to whether they assess the same abilities claimed by their developers in places where they are administered and whether the test interpretation in different EFL contexts is identical with that of the test developers'. Hence, in the context of Iranian EFL learners, the extent to which such tests turn out to be valid in testing language abilities needs close and thorough investigation.

Among the aforementioned tests, the FCE test method is quite new to the Iranian context and has been recently deployed in assessing the language abilities of Ph. D candidates at Isfahan University (Ph. D Entrance Exam 1385, TEFL Ph. D Entrance Exam1385). Hence the present study aims to investigate the way such a test is viewed by Iranian candidates and whether the items in the reading paper of the FCE measure the skills claimed by the FCE developing board (i.e. UCLES).

Application of the results of this study might even open windows to study the universal language tests introduced to Iranian educational context. It might also introduce new trends in the future university exams which have so far been confined to tests of TOEFL- like formats.

1.3. Significance of the Study

Since Reading comprehension is one of the basic skills in the process of second or foreign language learning, and many universal tests such as the FCE have this as one of their main sections, it seems that more research will shed light on what test items/sections assess in the context of EFL learners. Further, with reference to Iranian test takers, such studies seem to illuminate the way test items are completed by the test-takers and analyzed and/ or interpreted by language testing practitioners.

It has been strongly recommended by language testing researchers (e.g. Alderson 1990a, Weir *et al.* 2000) that, if reading skills are to be empirically operationalized, standardized tests of reading skills should be used. Consequently, universal tests such as the FCE, which are considered among the high-stakes tests claimed to be standardized, have to undergo quite strict and precise validation studies if they want to reduce the risks involved in the interpretations made upon their results.

Traditional approaches to construct validation appear to be inadequate for the purpose of examining the construct validity of reading tests. They largely ignore the processes that test-takers employ in taking tests, focusing on the content of the tests themselves and the products of whatever processes involved in taking a test (Anderson, Bachman, Perkins and Cohen 1991).

Recent thinking in educational measurement (e.g., Duran 1989 cited in Anderson *et al.*1991, Messick 1989, Farr, Pritchard and Smitten 1990) and in language testing (Grotjahn 1986, Bachman 1990) has begun to recognize the fact that when construct validity is investigated, the information about how test-takers go about processing test tasks should be considered.

The FCE is quite new to the Iranian EFL context, especially with regard to its test method facet. With the tendency observed in using test methods similar to that of the FCE, much of the test recently deployed in assessing the language proficiency of the Ph. D candidates at Isfahan University (Ph. D Entrance Exam 1385, TEFL Ph. D Entrance Exam1385), the results of this validation study might help stake holders in using clear interpretations of the test results. So there seems to be considerable space for work on the validity and appropriateness of universally standardized tests like the FCE in different contexts of language learning.

1.4. The Purpose of the Study

Recent studies on reading have moved away from a focus on product to investigating the reading process in order to better define the construct of reading comprehension (e.g. Farr, Pritchard, and Smitten, 1990; Pritchard, 1990). There is, however, the problem of specifying and adequately operationalizing the construct of reading comprehension (Rost 1993). According to Rost (1993) many studies have been published with a concentration on the analysis of reading comprehension skills. They have attempted to identify factors or components of reading comprehension and to develop tests that would allow a reliable and valid measurement of those factors. However, they are psychometrically oriented (e.g. Lennon 1962, MacGinitie 1973, Spache 1981, Groeben 1982, Johnston 1983).

In the same way, the present study attempts to investigate the reading comprehension skills; however, it deploys triangulated approaches of data collection to study the construct validity of the FCE reading paper with the focus on the skills used by the Iranian test-takers and expert judges. This study aims to tackle the problem of construct validity, i.e. the skills measured by the FCE reading paper, through the process of gathering different types of evidence: (1) EFL expert judgments (2) the EFL undergraduate test-takers' decisions on the skills assessed by the FCE reading items and finally (3) large scale test administration. The first two are qualitatively oriented and the third data collection approach is quantitatively oriented. The notion of 'identifiability of the FCE reading skills' will be dealt with, focusing on both qualitative and quantitative approaches of data collection.

1.5. Research Questions and Hypotheses

Based on the extant problems in the nature of reading construct and reading skills, the present study attempts to answer the following three questions:

RQ1- Do the majority of the Iranian EFL expert judges (half+1) agree on the skills measured by the items in the FCE reading paper?

RQ2- Do the majority of the Iranian EFL undergraduate test-takers (half+1) come to agreement on the skills measured by the items in the FCE reading paper?

RQ3- Do items in the FCE reading paper assess the same reading skills claimed by the test developers in the context of Iranian EFL undergraduates?

The above three research questions give rise to the following research hypotheses:

RH1-The majority of the Iranian EFL expert judges (half+1) agree on the skills measured by the items in the FCE reading paper.

RH2- The majority of the Iranian EFL undergraduate test-takers (half+1) agree on the skills measured by the items in the FCE reading paper.

RH3- Items in the FCE reading paper assess the same reading skills claimed by the test developers in the context of Iranian EFL undergraduates.

1.6. Definition of Key Terms

1- The First Certificate of English (FCE):

An examination produced by the University of Cambridge Local Examination Syndicate (UCLES), including a separate paper on reading with four parts, and 35 questions aimed at Level Three in the framework of the Association of Language Testers in Europe (ALTE)(FCE handbook 2001).

2- Construct:

A psychological concept which is derived from the theory of abilities (Alderson 2000:17). In this research it is the ability or skill (e.g. skimming or scanning) to be tested by a group of items in the FCE reading test.

3-Construct Validity:

Construct validity pertains to the meaningfulness and appropriateness of the interpretations that we make on the basis of test scores. To justify our interpretations, we need to provide evidence that the test score reflects the area(s) of language ability we want to measure, and very little else (Bachman & Palmer,

1996: 21). Bachman (1990) defines "construct validity" as the extent to which performance on tests is consistent with predictions that we make on the basis of the theory of abilities, or constructs.

4- Reading skills:

The abilities to process the information in the reading text are referred to as the reading skills (Alderson 2000).

5- Identifiability of Reading Skills:

Each individual skill in a list of reading skills can be taught, tested, and researched in isolation from others (e.g. Alderson 1990a, Weir 1993, Urquhart and Weir 1998, Alderson 2000,).

1.7. Outline of the study

The present study was organized in five chapters. In the first chapter, I have discussed the purpose (1.2) and the significance of the study (1.3). Research questions and hypotheses (1.5) have also been presented in the first chapter of the study.

The second chapter is a review on the well-reputed proficiency tests' development particularly the FCE (2.2). The FCE format, scoring, reliability and validity studies are then discussed in the consecutive sections (2.3). After the review of the related literature on the FCE (2.4), the concepts of validity, construct validity and the reading construct will be discussed in general (2.5_2.9). Finally the review on the reading skills is presented in the second part of chapter two (2.9).