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

Elahe Tavakoli
Higher Education Institute of Sobhe Sadegh, Isfahan, Iran
Email: Tavakoli.elahe@yahoo.com

Hossein Barati
University of Isfahan, Iran
Email: h.barati@gmail.com

Abstract—**Validity studies on language proficiency tests have attracted many researchers in the last decades. Most of such studies try to investigate skills assessed by the items in the tests, hence the construct validity of the test. The present study took the same approach and therefore aimed to investigate the construct validity of the reading paper of the First Certificate in English (FCE) in the Iranian EFL context. The research addressed 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 undergraduates come to agreement on the skills measured by the items in the FCE reading paper? (3) Does exploratory factor analysis support that the FCE reading paper assesses the same reading skills claimed by its developing board, i.e. UCLES, in the context of Iranian EFL undergraduates? The present study deployed triangulated approach in collecting and analyzing the data. It attempted to use both qualitative (i.e. experts' and test takers' judgment) and quantitative (i.e. Factor Analysis) approaches to address the above questions. The findings revealed that there was not a substantial agreement among expert judges on the skills claimed to be assessed by the items in the FCE reading paper nor could any significant agreement be observed amongst the test takers on the skills being measured by such items in the FCE reading paper. Finally, exploratory factor analysis revealed similar findings as those in the judgmental phase of the study. The individual items in the FCE reading paper did not confirm the claims by the test developers.**

Index Terms—**reading skills, construct, construct validity, inter-rater reliability, identifiability of reading skills**

I. INTRODUCTION

The notion of construct validity emerged out of efforts made in the early 1950s by the American Psychological Association to address the adequacy of psychological tests (Cronbach, 1988). Construct validity, as Bachman (1990, p. 255) defines, concerns "the extent to which performance on tests is consistent with predictions that we make on the basis of a theory of abilities, or constructs." Messick (1992), further, states that construct validity measures how much of something an individual displays or possesses. More specifically, Alderson (2000, p. 1) maintains "construct validity in reading as the ability we wish to test". It has also 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. This seems to be 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 reported evidence as to whether they assess the same skills claimed by their developers in their context of administration and whether score interpretation in various EFL contexts indicate similar results. Iran is one of the places where one of these universally administered tests (FCE) is frequently used. No study has been reported so far in the literature on the validity evidence of the FCE test in the Iranian EFL context. The present study, therefore, aims to investigate how such a test is viewed by Iranian candidates and whether the items in the reading paper of the FCE measure the same skills claimed by its developing board (i.e. UCLES).

University of Cambridge Local Examination Syndicate (UCLES) has devised a set of five examinations at 5 levels of proficiency. The FCE is at level 3 within the series and corresponds with what is referred to as an intermediate stage of proficiency (FCE handbook 2001). The FCE was originally offered to the field of language testing and teaching as the "Lower Certificate of Proficiency" in 1939. In 1974, it was renamed as the "First Certificate in English" by UCLES. It is now one of the most widely taken Cambridge EFL examinations and its annual worldwide candidature is in excess of 250,000 (FCE handbook, 2001). The FCE comprises 5 different sub-tests/papers: (I) Reading, (II) Writing (III) Use of English (IV) Listening (V) Speaking. There are 4 sections within paper 1- the reading paper (the paper under focus of this research) with a total of 35 reading comprehension questions. The FCE handbook (UCLES 2001) claims that the primary concern for the test reading paper is to focus on various reading skills such as identifying the main points, details, opinion and gist, deducing meaning and identifying text structure through a gapped text. The first subtest in the

FCE reading paper, therefore, is believed to measure the test takers' ability to recognize main idea, of the text at paragraph level. In this section, the candidates are asked to match headings from a list with individual paragraphs in a long text. The second is claimed to assess detailed as well as global understanding of the text, the ability to infer meaning and lexical reference. Part 3 also is suggested by the developing board to assess the test takers' understanding of how texts are structured. And finally the forth part requires the test takers' to locate specific information in the text comprising individual paragraphs. The task in this last part requires the candidates to scan and skim the text and while doing that, look for the appropriate paragraphs in the text to match them with the related statements (FCE handbook 2001).

Bachman, Ryan and Choi (1995), investigated the construct validity of a set of examinations including the FCE and TOEFL. They focused on the abilities measured in each test battery in two ways: (i) the qualitative content analysis of the two tests, including the specific language abilities and the type of the test task employed and (ii) the quantitative investigation of patterns of relationship in examinees' performance on the tests, both at the level of total test scores, and, where appropriate, at the item level.

Furthermore, in order to compare the abilities measured by the FCE and the TOEFL, Bachman and his colleagues (1995) examined the correlation coefficient between (i) the scale scores for the five versions of the FCE paper, (ii) the standardized scores for ETS measures, and (iii) all the thirteen of these measures. For the first and the second stages, exploratory factor analysis was employed and the results indicated that:

In the case of the FCE papers, all loaded most heavily on the general factor, which accounted for slightly over half of the total variance in the test… This suggests that the FCE papers all tend to measure a single language ability, with specific abilities measured by paper 3 (use of English) and 5 (speaking), to a lesser degree by papers 1 (Reading), 2 (Writing) and 4 (Listening) (Bachman *et al*., 1995, p. 65).

Moreover, the correlation coefficient between all the thirteen measures of the FCE and the ETS tests, together with exploratory factor analysis indicated that:

Although the scree test suggested that only two or three factors should be extracted, the parallel factors should be extracted; the parallel factors extracted indicated five. … The first factor can be interpreted as 'SPEAK' factor, the second on'ETS structure', 'reading' factor and the fourth a 'listening and integrative speaking' factor (For more information see Bachman *et al*., 1995, pp.69-72).

Finally Bachman and his colleagues concluded that "the large portions of variance accounted for by the general factors in our analysis suggested that the FCE papers and the ETS tests appear to measure, to a large degree, the same common aspect of the language proficiency of the subjects in our sample."

The purpose of this study is therefore to scrutinize the construct validity of the FCE reading paper within the context of Iranian test takers. The study focuses on the skills claimed by the FCE developing board as assessed by the test items and the EFL expert judges and what the individual test takers believed to be assessed by the FCE reading paper test items. Finally a large scale test administration will be conducted to examine if exploratory factor analysis confirms what the FCE developing board states in relation to skills under focus by the individual items in the FCE reading paper. The notion of 'identifiability of the FCE reading skills' will be dealt with through both qualitative and quantitative approaches of data collection. The study, therefore, addresses the following research questions:

1. Do the Iranian EFL expert judges come to a substantial agreement on the skills measured by the items in the FCE reading paper?

2. Do the Iranian EFL undergraduate test takers come to agreement on the skills measured by the items in the FCE reading paper?

3. Do items in the FCE reading paper assess the same reading skills claimed by the test developing board, i.e. UCLES, in the context of Iranian EFL undergraduates?

II. METHOD

A. Participants

The construct of the FCE reading paper was validated in three stages:

(1) against the EFL expert judgment

(2) against the EFL undergraduate test takers' decision

(3) through large scale test administration and running exploratory factor analysis on the scores

In order to address the first research question, 6 university lecturers were asked to complete the FCE reading paper as EFL learners. They were selected from the Isfahan state University and Islamic Azad University of Najafabad. They had teaching experience for more than 5 years. Three of the lecturers held Ph. D in teaching English; the other three held Master's degree in TEFL or translation studies who had similarly more than 5 years of teaching experience to EFL learners of the same characteristics as those participating in the study.

As for the third stage, the quantitative phase of the study, the FCE reading paper was administered among 150 Iranian Junior and senior students studying in the third and fourth year of their university. The participants in this stage had similar characteristics with those of the second stage.

B. Materials

1. The FCE reading paper of the UCLES' back issues

The 35-item reading paper (March 2002) was used in the present study. The scoring procedure followed the guideline provided by the FCE handbook (UCLES 2001). The maximum test score for the FCE reading paper was 56. The time allowed for the completion of the reading paper, as determined by UCLES and confirmed in a pilot study, was 75 minutes.

2. The EFL taxonomy of reading skills

The second material deployed in this research was a taxonomy of EFL reading skills (Barati 2005) (appendix A). This taxonomy was categorized based on the levels in the Weir (1997) checklist. The present study used the same taxonomy (Barati 2005) since it was thorough and had been validated qualitatively against the experts' judgments.

C. Procedure

The procedure of the study comprised three phases. First, 6 university lecturers were asked to contribute to this study as expert judges. They were especially asked to take the FCE reading paper as the EFL learners and decide on the correspondence between individual test item in the FCE reading paper and the skills in the EFL taxonomy of reading skills (Barati 2005). The judges were allowed to take the test home and complete the tasks required, thus there was no cooperation among the judges.

Second, 6 undergraduate students were asked to complete the FCE reading paper. Similar to the previous group, these students were provided with the FCE reading paper and the taxonomy of skills (Barati 2005). They were required to match the skill they thought was assessed by each test item in the FCE reading paper on the EFL reading taxonomy of reading skills.

Finally, the FCE reading paper was administered to 150 male and female undergraduates. The test was administered to 5 different groups. In all the test sessions, attempts were made to follow the same procedures.

In every test session and before distributing the test, the aims of the researcher were briefly outlined. The participants were informed of the significance of the proficiency tests including the FCE. To give the test takers more motivation, they were also informed that the results of their performance would be presented to them later to make them aware of their own level of reading ability.

III. DATA ANALYSIS

To answer the first and the second research questions, the frequency of the experts' and learners' judgments over skills/item correspondence was investigated. Also inter rater reliability of experts' judgment and learners' decisions was employed to make sure there was no significant discrepancy in the decisions made.

To quantitatively analyze the data, exploratory factor analysis was run on the scores obtained through large scale test administration. In so doing, first, descriptive data analysis was conducted to examine the normal distribution of the variables across the FCE reading paper. Secondly, the reliability estimates of the FCE reading total and its subparts were calculated using Cronbach's Alpha. Finally the reading skills were identified by running exploratory factor analysis. The assumption was that items which loaded on the same component or factor were testing the same skill/s (Alderson 2000).

Normality and the Reliability

The normality of variables, expressed in terms of'skewedness'and 'kurtosis', is significant to any quantitative data analysis (Bradley 1982). The analysis of normality at item level for the FCE reading paper indicated most of the items within the recommended normal range (+2 and -2) yet few items (2,5,6,10,19,23,32) with the normality above 2.

Following the idea that without the reliability the validity of the test is threatened (e.g. Alderson 1990a, Chapelle 1999), Cornbach's Alpha Coefficient was estimated for both the total FCE reading paper and each of its subtests since each subtest in the FCE reading paper is claimed to assess a different set of traits/skills (Table 1).

TABLE 1:
RELIABILITY ESTIMATES, THE FCE READING PAPER: ITS TOTAL AND SUB-TESTS AND TOTAL

Subtest	No. of items	Items	Alpha
1	7	1-7	.694
2	7	8-14	.331
3	7	15-21	.640
4	14	22-35	.857
Total FCE reading paper	35		.839

Cronbach's Alpha Coefficient of .83 showed a reasonably high internal consistency. However, two of the FCE reading sub-tests (1and 3) met a moderate alpha coefficient and the second subtest (the subtest with multiple choice items) showed a low alpha coefficient index (Cronbach 1951).

IV. RESULTS

A. EFL Expert Judges and the Identifiability of the FCE Reading Skills

The responses of the expert judges' over skill/item correspondence were put into more quantitative analysis by conducting the inter rater reliability between every pair of the judges' decisions. The results revealed that there was not a significant agreement amongst the majority of the expert judges (i.e. more than half of the paired expert judges). Only 3 of the paired judges (out of 15 pairs) showed an agreement on the skills to be assessed by the FCE reading paper (Table 2). Table 2 shows the inter-rater reliabilities at a moderate level (i.e. correlations between .50 and .75).

TABLE 2:
EXPERT JUDGES' INTER RATER RELIABILITY

Experts	Inter-rater reliability
Judge 1 &2	.51
Judge 2 &4	.68
Judge 3 & 5	.57

The present study also adopted frequency measures to analyse the agreement of the EFL expert judges on the individual FCE items. The frequency of the experts' decisions on the skill assessed by each item also indicated that the majority of the judges (half+1) did not come to an agreement on the skill to be assessed by the majority of the items of the FCE reading paper. Only 6 items of the FCE reading paper (items 4, 15, 16, 17, 18 and 21) (3) could show an agreement of the majority of the expert judges.

TABLE 3:
FREQUENCY OF THE EFL EXPERTS' DECISIONS ON THE ITEM/ SKILL CORRESPONDENCE

FCE items	The selected skill from the EFL taxonomy	The frequency of the skill	Percentage
item 4 (part 1)	3 (getting the main idea)	4	66.7%
item 15 (part 3)	12 (attempting to find the relationship between ideas in the passage)	5	83%
item 16 (part 3)	12	4	66.7%
item 17 (part 3)	12	5	83%
item 18 (part 3)	12	4	66.7%
item 21 (part 3)	12	4	66%

B. EFL Test Takers and the Identifiability of the FCE Reading Skill

In order to obtain EFL learners' opinions on the skills being tested by the FCE reading paper, the same procedures as those used for the expert judges were applied. This phase of data analysis revealed that there were no pairs of the test-takers with a considerable inter-rater reliability. The frequency of the test-takers' decisions on the skill assessed by each item indicated that the majority of them (66.7%) agreed on the skill identified (skill 12) only for item 18 of the FCE reading paper. In other words, the majority of the EFL test-takers could only come to an agreement on 2.8% of the items in the FCE reading paper (i.e. 1 out of 35 items in the FCE reading paper) and therefore they showed no significant agreement on the skills measured by the FCE reading skills.

The frequency of test-takers' decisions on the individual FCE reading item could only show few similarities with the findings of the expert judges. Both groups: the EFL experts and the EFL test-takers agreed on the skill 22 to be assessed by item 11 (in part 2) and skill 12 for item 18 yet, this agreement could not be significant considering all the test items. In the same way, neither the inter-rater reliability estimates nor the frequency measures showed the majority agreement between the EFL expert judges or the EFL test takers.

C. Quantitative Factor Analysis and the Identifiability of the FCE Reading Skills

It is commonplace in research on reading skills that identifiability of skills is addressed by applying factor analysis (FA) on the data (e.g. Rost, 1993, Alderson, 2000). The idea is that items loaded on the same factor, assess the same skills. Before conducting FA; however, the test items were inspected for the correlation matrix and the correlation coefficients of .3 or above. The results revealed that there were some correlation coefficients of .3 or above. Furthermore, the test of factorability of data (i.e., The Kaiser Meyer Oklin) exceeded the recommended value of .6 (Kaiser, 1974) and the Barlett's test of sphericity (Barlett, 1954) reached statistical significance (Table 4), therefore the factorability of the correlation matrix is supported.

TABLE 4:
TEST OF FACTORABILITY OF DATA, FCE (KMO AND BARTLETT'S TEST)

| Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.797
Bartlett's Test of Sphericity	Approx. Chi-Square	1445.564
	df	595
	Sig.	.000

The application of factor analysis at this stage resulted in 11 components with eigenvalues of 1 or above, explaining 79.3% of the total variance (Table5). As it is indicated in Table 5, 21 items of the FCE reading paper were loaded on the first factor. The remaining 14 FCE reading items did not load with a significant variance.

TABLE 5:
COMPONENT MATRIX FOR THE FCE READING PAPER

	Component										
	1	2	3	4	5	6	7	8	9	10	11
FCE35	.684										
FCE31	.660										
FCE25	.653										
FCE27	.646										
FCE34	.627										
FCE3	.612									.342	
FCE33	.599										
FCE23	.582										
FCE24	.582										
FCE22	.552										
FCE12	.530			.308							
FCE1	.521	-.350									
FCE26	.511										
FCE32	.504									-.337	
FCE7	.466		.344					.458			
FCE29	.463								-.415		
FCE4	.455		.348								
FCE30	.443								-.396		
FCE5	.425				.352						
FCE16	.353	.549									
FCE21	.335	.478									
ECE18		.437	.425								
FCE19	.308	.433									
FCE20		.454	.521								
FCE9	.323		-.439			-.339					
FCE6	.377		.382								
FCE8				.690							
FCE10				.560			-.451			.354	
FCE11					-.521	.384	.375				
FCE15	.392		-.319	.455							
FCE17	.342	.304		.424							
FCE2					.377	.500			.329		.315
FCE14		-.331		.377		-.416					
FCE13	.373						.424	-.412			
FCE28	.383								-.397		

Extraction Method: Principal Component Analysis, 11 components extracted

The Screeplot, similarly, revealed that there was a clear break after the first component. This factor/component was responsible for 20.90% of the total variance. Thus no further rotation was conducted on the data.

Scree Plot
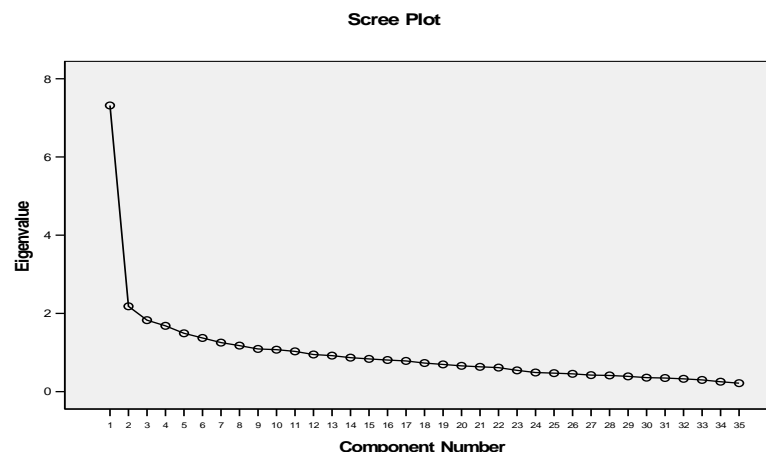
The aforementioned findings suggested that the FCE reading paper is not such a robust measure for distinguishing among skills. It, therefore, did not provide a clear picture of assessing the skills for which the sub-tests were developed. In other words, the items in the FCE did not identify among the skills in the Iranian EFL context. Thus, the findings obtained from running FA did not support the identifiability of the reading skills in the Iranian context and rejected the third hypothesis.

V. DISCUSSION

The findings of this study did not show a significant agreement among the majority of the Iranian expert judges nor the EFL test takers on the skills to be assessed by the items in the FCE reading paper. The qualitative judgmental findings of this study were, therefore, in line with what many other researchers have concluded (e.g. Alderson and Lukmani 1989, Alderson 1990a and 1990b). The first two studies had applied the same two-phased methodology, judgmental and empirical, to address identifiability of skills. However, despite Alderson and Lukmani (1989), Alderson (1990a) conducted his research on two standardized tests, TEEP and ELTS). Alderson (1990b), recognizing the inadequacy of traditional methodologies, attempted to use think aloud and retrospection and identify what really happened inside the test-takers' head as they responded the reading comprehension items. His conclusion resembled those of his previous two studies and the present study. They could not prove the existence of separate skills.

The findings of the first research question were against Lumley (1993) and Barati (2005). Unlike the present study which focused on a standardized high-stake reading test, having been developed by a specialized examination board, Barati's (2005) test was partly adopted from Alderson (2000), and Weir (1993) and partly developed by himself. Lumley (1993), however, initiated his project with the test items and then developed the reading skill to describe them. Both studies clearly supported the judges' agreement on the skills to be measured by their reading items through judgmental approach.

The assumption was that if the test-takers could agree on identifying individual skill associated with each item or the total FCE reading paper, the skills would be seen as identifiable and hence the reading paper could be identified as a valid test in Iranian EFL undergraduate context. Alderson (2000, p. 97) interpreted test takers' different approaches in taking the test as a problem to the validity of the test.

If different test takers respond differently to an item, and yet get the item correct, there is a real problem in determining what the item and the test is testing The validity of a test relates to the interpretation of the correct responses to items, so what matters is not what the test constructors believe an item to be testing, but which responses are considered correct by the test-takers, and what process underlies them (p. 97).

The inter-rater reliability coefficient estimated between 15 pairs of the EFL undergraduate test-takers showed no agreement among them on the skills they employed when completing the FCE reading paper. The results of the second research question were in line with what Li (1992) concluded (Cited in Alderson, 2000). Li (1992) asked a group of test-takers to introspect on the skills assessed by items in his reading test. He concluded (i) seldom did the test-takers report using one skill alone in answering test items (ii) when the skills used corresponded to the test constructor's intentions, the students did not necessarily get the answer correct and (iii) students answered correctly whilst using skills that the test constructor had not identified. Thus items in Li's reading comprehension test did not necessarily assess what the constructor claimed his study showed no agreement on the skill the test-takers employed to answer individual items.

Following the common assumption that reading skills are identifiable by applying factor analysis (FA) (e.g. Rost, 1993, Alderson, 2000), the data obtained from the FCE administration was put into exploratory Factor Analysis. The results revealed empirically that most of the items in the FCE reading paper (21 items out of 35) loaded on one component (variable) and only measured one skill, in other words, 60% of the FCE reading items loaded strongly on

one component and the FCE reading paper was shown to be unitary in its construct. The unitary nature of the FCE reading paper in the Iranian EFL context contradicts UCLES' claims about the various skills and in turn is contradictory with different interpretations that are made based on the test results in the Iranian context.

The fact that the FCE items in this study didn't assess different skills does not seem surprising. A few other studies provided evidence for the unitary nature of the reading construct in the EFL tests such as TEEP and IELTS (Alderson, 1990a, 1990b). However the question is: how is it possible for a standardized test which is administered universally and its scores are interpreted on the basis of the specifications of its items and the skills (the FCE handbook 2001) not to assess what its developers claim to?

The application of Factor Analysis on the data from this study indicated that the standardization of the FCE reading paper does not guarantee 'universal' interpretation of its scores. A test may be standard in relation to a certain group of test takers with a specific language background and system of education but not in relation to another group of test takers with different language background and system of education. This seems in concert with Bachman, Davidson, Ryan, and Choi's (1995) suggestion in relation to their comparative study of the tests of Educational Testing Service (ETS) and UCLES specially TOEFL and the FCE.

The quantitative findings of Bachman *et al.* (1995) showed a significant difference in their test takers' performance due to their preparation or familiarity with the tests. The US test-takers in their investigation performed higher on the TOEFL whereas the UK test-takers performed higher on the FCE and CPE.

The quantitative findings of the present study supported Bachman *et al.* (1995) findings. Although their study did not focus on the validation of the FCE in particular, they came to over half of the total observed variance in the test loaded heavily on one general factor. In other words, they report that the FCE papers- involving the reading paper- tended to measure one single language ability.

Language proficiency tests tap heterogeneous populations as their test takers' target population. The FCE is not an exception, although it is claimed to be an intermediate test. This is what Woods (1993) refers to as the heterogeneity of the FCE tasks and the candidates. He suggests:

Communicative language tests- such as the FCE- which are task based, may be heterogeneous in two ways: (i) the tasks tap a broad range of language skills; and (ii) the candidates bring very different profiles of skills to bear, which may be taken to represent equally valid expressions of ability (Woods 1993 cited in Weir 2005: 32).

According to UCLES (2004), the consequence of these heterogeneities would be that items take longer time to complete, and hence fewer items can be accommodated within practical time constraints. This may bring about a small reduction in the estimated reliability of the FCE reading paper when compared with tests such as TOEFL and IELTS (UCLES, 2004).

However, despite the aforementioned factors which, according to UCLES (2004), bring about the reduction in estimated reliability, the total FCE reliability estimates in this study tends to be reasonably high (r=.86). The reason for the high reliability of the test could be attributed to the large scale administration of the test to 150 participants in the study.

The important point to make is that, the FCE reading paper like many other Cambridge examinations is claimed to be heterogeneous in the tasks and the skills although they might not show a high index of reliability (Woods, 1993). The FCE reading paper in the present study could; however, show a reasonably high estimate of reliability, though it could not distinguish various skills in the Iranian EFL context.

VI. CONCLUSION

This study, adopting qualitative and quantitative approaches, has examined the identifiability of the FCE reading skills with a focus on the construct validity of the reading paper. The quantitative phase of the study involved the analysis of 150 adult EFL undergraduate test-takers' responses on the FCE reading paper and the application of FA. The qualitative phase of the study involved EFL expert judges' and the EFL test takers' decisions on the correspondence between the reading skills in the EFL taxonomy and the individual items in the FCE reading paper.

The key findings of the present study, as discussed in 2.7 above, indicated that the FCE reading skills were not identifiable whether quantitatively via FA or qualitatively by EFL experts and learners. Hence the findings are against UCLES' claims, that the FCE reading paper assesses different skills. In other words, the findings of the present study degraded the construct validity of the FCE reading paper in the Iranian context and rejected all the three research hypotheses.

The judgmental phase of the present study could support the findings of the quantitative phase (FA). In the qualitative phase of the study, the majority of the EFL expert judges and the test-takers did not agree on the skills to be assessed by the FCE reading items. In other words, the FCE reading skills were not similarly identified by the expert judges and the test-takers to be assessing heterogeneous list of skills. Neither were the skills identified when Factor Analysis was run. The results of the Factor Analysis did not show the items in the FCE reading paper to be loaded on different components. Neither the qualitative nor the quantitative phase of this study could identify different skills for the construct of the FCE reading paper, hence, the construct of the FCE reading paper was known to be unitary in the context of Iranian EFL learners

APPENDIX A: THE EFL TAXONOMY OF READING SKILLS ACCORDING TO LEVELS IN WEIR'S (1997) CHECKLIST, (BARATI 2005)

Level A: Reading Expeditiously for Global Comprehension
   1. Looking at the topic, title and sometimes the first sentence of the paragraph to skim the text
   2. Reading the first and the last sentences of the paragraph
   3. getting the main idea
   4. attempting to find the gist of the passage

Level B: Reading Expeditiously for Local Comprehension
   5. scanning
   6. looking for specific information

Level C: Reading Carefully for Global Comprehension
   7. reading everything carefully
   8. paying attention to components of a paragraph: topic sentence, supporting sentences and concluding sentences
   9. reading details
   10. evaluating the text in order to find its implications
   11. being able to answer the questions which concern key information from text (without which the comprehension of the text would be hampered)
   12. attempting to find the relationships between ideas in the passage
   13. being able to distinguish between major and minor points in the text
   14. understanding implications
   15. understanding facts and major points
   16. inferential analysis of text
   17. being able to answer detailed questions about inferences and reasoning

Level D: Reading carefully for Local Comprehension
   18. Guessing the meaning of unfamiliar words
   19. knowing how to attack lexical difficulties (finding the stems, word etymology, using context, etc)
   20. paying attention to sentence structure to guess the meaning of unknown words
   21. paying attention to context to guess the meaning of unknown words
   22. identifying the syntactic relationship between the items in a text to comprehend it better (the pronoun and its referents, conjunctions and other elements)
   23. understanding the key words, key structures
   24. structural analysis of text
   25. identifying words synonyms and antonyms
   26. knowing the vocabulary and the connection with the sentence
   27. knowing grammatical terms, verbs, adjectives,… to understand the text
   28. realizing what the words do in the text
   29. knowing the grammatical structure of the text to assist understanding
   30. being able to answer detailed questions about words and phrases
   31. being able to use the structure of the word to get its meaning
   32. decomposing the word

REFERENCES

[1] Alderson, J. C., & Lukmani, Y. (1989). Cognition and reading: Cognitive levels as embodied in test questions: *Reading in a foreign language, 5(2),* 253-270.
[2] Alderson, J. C. (1990a). Testing reading comprehension skills. *Reading in a foreign language*, 6(2), 425-438.
[3] Alderson, J.C. (1990b). Testing reading comprehension skills. Getting students to talk about a reading test (A pilot Study). *Reading in a foreign language*, *7 (1)*, 465-503.
[4] Alderson, J. C. (2000). Assessing Reading. Cambridge: Cambridge University Press.
[5] Bachman, L. F. (1990). Fundamental considerations in language testing. Oxford: Oxford University Press.
[6] Bachman, L. F., Davidson, F., Ryan, K., & Choi, I. C. (1995). An investigation into the comparability of two tests of English as a foreign language: The Cambridge TOEFL comparability study: Final Report. Cambridge: Cambridge University Press.
[7] Barati, H. (2005). Test taking strategies and the assessment of reading skills: An approach to construct validation. Unpublished Ph. D Thesis. University of Bristol.
[8] Barlett, M. S. (1954). A note on the multiplying factors for various chi-square approximations. *Journal of the Royal Statistical Society*,16 (2), 296-298.
[9] Chapelle, C. (1999). Validity in language assessment. *Annual Review of Applied Linguistics,* 19, 254-72.
[10] Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika.* 16, 292-334.
[11] Cronbach, L. J. (1988). Construct Validation after thirty years. in Robert L. Linn (ed.). *Intelligence, Measurement, Theory, and Public Policy*. Urbana, Ill: University of Illinois Press.
[12] Li, W. (1992). What is a test testing? An investigation of the agreement between students' test taking processes and test constructors' presumptions. In J. C. Alderson. (2000). *Assessing Reading*. Cambridge: Cambridge University Press.

[13] Lumley, T. (1993). The notion of sub-skills in reading comprehension tests: an EAP example. *Language Testing*, *10(3)*, 211-234.
[14] Messick, S. A. (1992). Validity of test interpretation and use. *Encylopedia of educational research (6th edition).* New York: Macmillan.
[15] Rost, D. H. (1993). Assessing the different components of reading comprehension: Fact or Fiction. *Language Testing*, *10(1),* 79-92.
[16] UCLES. (2001). First Certificate in English (FCE) handbook. Cambridge: UCLES.
[17] UCLES. (March 2002). First Certificate in English (Past Examination Papers). Cambridge: University of Local Examination Syndicate.
[18] UCLES. (2004). EFL research notes1. Cambridge: Cambridge University Press.
[19] Weir, C. J. (1993).Understanding and developing language tests. London: Prentice Hall.
[20] Weir, C.J. & porter, D. (1994).The multi-divisibility or unitary nature of reading: The language tester between Scylla and Charybdis. *Reading in a foreign language*, 10(2), 1-19.
[21] Weir, C., J. (1997). The testing of reading in a second language. In C. Clapham., & D. Corson., (eds.) Encyclopedia of language and education. *Language testing and assessment*, 7, 39-49.
[22] Weir, C.J., Huizhong, Y., & Yan, J. (2000*)*. An Empirical Investigation of the Componentiality of L2 Reading in English for Academic Purposes: Studies in language testing 12. Cambridge: Cambridge University Press.
[23] Weir, C.J. (2005). Language testing and validation**.** UK**:** Macmillan.
[24] Woods, R. (1993). Assessment and Testing. In C.J. Weir., (2005). *Language testing and validation***.** U.K**:** Macmillan.

Elahe Tavakoli was born in Isfahan, Iran in 1980. She received her M.A. degree in TEFL from University of Isfahan in 2007.

She is currently a lecturer in the higher Education Institute of Sobhe Sadegh, Isfahan. Her research interests include Reading Skills and strategies, and Testing.

Hossein Barati works in the English department, University of Isfahan, where he is assistant Professor of Applied Linguistics. He has a PhD in 'Language Testing' from University of Bristol and has been involved in research in language testing, programme evaluation, reading strategies, and classroom discourse. Before becoming a university lecturer in 1992, he worked as a language teacher at Ministry of Education, Isfahan, Iran. He has published in areas of language testing and assessment, and language programme evaluation. He is a member of a research project on *Developing a new model of teaching to adult English non-majors at University of Isfahan*, funded by University of Isfahan.

He has also co-authored some publications on DIF in Iranian National University Entrance Exam (INUEE), 2007; Linguistic constraints and language teaching, 2008; and Cultural differences in EFL performance on cloze tests, 2010. Further a text-book *A step forward in English for intermediate EFL learners* is among his latest publications.

Hossein is currently working on '*New plans for Teaching English to Iranian adult non-majors'*, '*Teaching English to Iranian young learners*', and '*The consequential validity of high stakes tests in the Iranian context'*.