Journal of
Language and Translation
Volume 5, Number 2(10), (pp.43-59), 2015

Evaluation of the Iranian State University EFL Entrance Examination Test (UEEET)

Gholam-Reza Abbasian[1*], Hedieh Nassirian [2]
[1] Imam Ali & IA (South Tehran Branch) Universities
[2] Islamic Azad University (South Tehran Branch)

Received: 5 March, 2015 Accepted: 13 August, 2015

Abstract
The motive behind this study was to evaluate the Iranian State University EFL Entrance Examination Test (UEEET) based on Bachman and Palmer's (1996) 'Usefulness Six-faceted Model 'accommodating Reliability, Validity, Impact, Interactiveness, Authenticity, and Practicality. To do so, thirty professors and one-hundred EFL freshmen were selected randomly from five universities. A questionnaire consisting of thirty likert-scale items and six open-ended questions were developed based on Bachman and Palmer's (1996) framework. They were administered to both groups of the participants along with the 2013 sample of the Iranian State UEEET. In addition to the questionnaire, a structured interview consisting of ten questions was also run for the purpose of triangulation of the data. Based on the findings, both groups evaluated the UEEET mainly less reliable, less practical and entailing negative impacts, while they held positive views towards its validity, authenticity, and interactiveness. Certain contradictions in the results leave the space open for further investigations, of course. Overall, the UEEET is evaluated as a test enjoying an acceptable degree of the Usefulness criteria.

Keywords: Iranian State University EFL Entrance Examination, Test Usefulness, Test Evaluation

INTRODUCTION

It is worth initiating the introduction with a quotation from Fulcher (2010, p.277) when he says: "The effects of the use of language tests are the measure of the meaning of the test in practice. If the test has been well designed, with purpose and effects in mind, we might expect to see many positive practical effects for most stakeholders". This quotation equates test effectiveness, in fact, with test mindfulness, which is a clear-cut index for the necessity of designing a more effective, purposeful, and well-designed test. The design and construction of a good test, however, is a skill that requires not only knowledge of the field and clear view of the desired outcomes, but also psychological understanding of pupils. Moreover, when the objective of conducting a test is to compare examinees' performance, it is then necessary to provide equal conditions for all group members so that they can show their knowledge and abilities.

Reiterating on the role of teachers and instructors in test and test item development, Farhady (2006) claimed that not all qualified teachers can write appropriate test items. Thus, writing test

*Corresponding Author's Email: gabbasian@gmail.com

items is a difficult job because even a single inappropriate item might change the outcomes of the exam, and student's failure might be the result of poor test construction rather than his/her lack of knowledge. Then, test design, development and administration get to be of prime significance in this process of decision-making.

Significance of a test lies in two macro issues: complexity of decisions made thereby and multiplicity of the nature of the test by itself. As far as the former is concerned, the decisions made by a language test according to Bachman (1981, as cited in Bachman, 1990, p, 58) are of two types: decisions about individuals called micro – decisions and decisions about the program identified as macro-decisions. The first types range from the selection, progress check, etc to decisions about individual teacher, while the latter revolves around the areas of program. The second area of test significance; the multiplicity of its nature, which is the focus of and rationale behind this study, is rather complex and highly interwoven with the first issue; decision type. In other words, theses two issues are not only mutually exclusive rather mutually complementary and interdependent. It is rationalized on the ground that nature of test design requires that it can accommodate such a multiplicity, and complexity of decision by itself necessitates that the instrument be multi-faceted. All these are integrated in a term called 'Usefulness' coined most probably by Bachman and Palmer (1996). Of course, language tests enjoy research value as well in that they can be used as an illuminating device on the nature of language proficiency, language processing, language acquisition, language attrition, and so on (Bachman, 1990)

As an evidence of the interdependence of the two mentioned issues (i.e., *complexity of decisions made and multiplicity of the nature of the test by itself),* Bachman and Palmer (1996, p. 17) claim that " the most important considerations in design and development of a language test is the use for which it is intended, so that the most important quality of a test is its usefulness". Then, the test 'usefulness' "provides a kind of metric by which we can evaluate not only the tests that we develop and use, but all aspects of test development and use" (ibid). For them, 'a model of test usefulness' is used as a measure of quality control both through the entire test development and use processes (Bachman and Palmer, 1996). Such a model includes six qualities schematized as follows:

Test Usefulness= Reliability+ Construct validity+ Authenticity+ Interactiveness+ Impact+ Practicality

Adapted from Bachman & Palmer (1996, p.18)

Language test development, use and its evaluation have been commonly approached narrowly since only too few aspects have been identified as the test characteristics. For example, Farhady, Ja'farpour and Birjandi (2006) characterize a good language test possessing acceptable indices of validity, reliability and practicality. Similarly, Bachman (1990) focuses on validity and reliability and Brown and Hudson (2003) expand the areas to reliability, validity, dependability, and uni-dimensionality. To the best of the researchers' knowledge, rarely have language tests been approached from multiple perspectives, contrary to the parallel developments in the theories and methods of language teaching and learning (e.g., CC, CLT, NA, TBLT, etc).

In line with the developments in the theories of language teaching and learning, specially in light of Communicative Language Teaching (CLT) based on which social and realistic use of language teaching received prime significance, Messick (1980 as cited in Bachman, 1990) introduces the notion of 'validity as a unitary concept' followed by Bachman's (1990) notion of 'consequential or ethical basis of validity'; all in a bid to relate the type of decisions made by a test to target language use (TLU). Further developments in the field of language test development and use led to the expansion of the dimensions of test characteristics and evaluation such that Bachman and Palmer (1996, p. 13) offered a new philosophy of language testing as follows:

1. *Relate language testing to language teaching and use.*
2. *Design your tests so as to encourage and enable test takers to, perform at their highest level of quality.*
3. *Build considerations of fairness into test design.*
4. *Humanize the testing process: seek ways in which to involve test takers more directly in the testing process; treat test takers as responsible individuals; provide them with as complete information about the entire testing procedure as possible.*
5. *Demand accountability for test use; hold yourself, as well as any others who use your test, accountable for the way your test is used.*
6. *Recognize that decisions based on test scores are fraught with dilemma, and that there are no universal answers to these.*

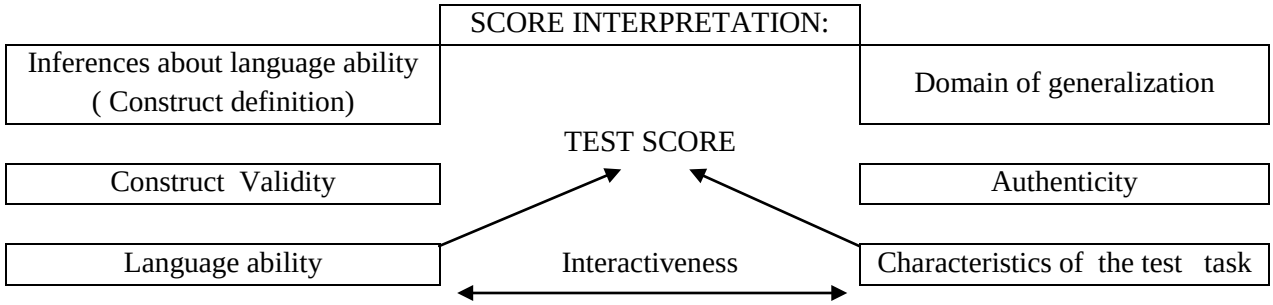
Based on this philosophy, test design, use and evaluation have been revolutionized such that they have offered their test '*Usefulness Model'*, used as a practical framework. Major dimensions of test usefulness according to Bachman and Palmer (1996) are then briefly addressed as follows:

Reliability

Defined as consistency of measurement, reliability covers and should be evaluated based on a range of parameters such as variation of test scores in terms of administration, test rubrics and forms, input and response type and characteristics and relevance, etc.

Construct validity

Revolving around the 'meaningfulness and appropriateness of the interpretations made based on test scores', construct validity covers definition of the target construct and the pertinent test task characteristics. Such aspects are illustrated in the following figure (Bachman and Palmer, 1996, p.22):



Construct validity is measured based on the extent of the richness and kinds of evidence possible to be gathered and offered as to the construct definition and the domain of generalizations in terms of the correspondence between the features of TLU tasks and test tasks. Furthermore, construct validity criteria cover areas such as transparent definition of the target construct, construct-test purpose relevance, test task and construct match, test scoring and scores match to the construct and accessibility of desired interpre tations on the desired ability, biasnesses originating from the test setting, rubric, input and expected responses effects on performance, and relationship between input and response causing differences in performance.

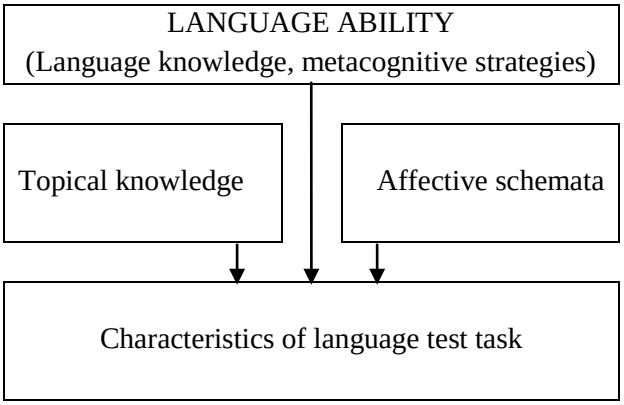
Authenticity

In line with the construct validity, authenticity is defined as 'the degree of correspondence of the characteristics of a given language test task to the features of a TLU task as represented by Bachman and Palmer (1996, p. 23):

Moreover, Bachman (1990) holds that " real life performance as a criterion for authenticity" ……the real life (RL) approach to defining authenticity essential considers the extent to which test performance replicates some specified non-test language performance" (p. 301); analogous to Bachman and Palmer's TLU performance. Authenticity is measured based on two ways including task characteristics and expected perceptions on the part of test takers and users. These criteria are realized based on the characteristics of TLU domain and match between test task and TLU tasks.

Interactiveness

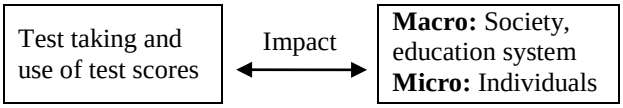
Interactiveness refers to the extent and type of engaging the characteristics of individual test taker in doing a test task. Such characteristics are better illustrated in the following figure (Bachman and Palmer 1996,p.26):



Then, test interactivenesss includes measures and areas of test taker's language ability, topical knowledge, affective schemata and task characteristics that each of which covers many dimensions as elaborated by Bachman and Palmer (1996) and documented in the pertinent Questionnaire designed for the purpose of this study.

Impact

Conservatively taken as interchangeable with Bachman's (1990) consequential validity, impact of a test refers to its 'impact on society and educational systems and upon the individuals within those systems' it is represented in the following figure: (Bachman and Palmer 1996, p.30)



Impact of a test can be measured based on many questions concerning the test takers, teachers, educational systems, and society. So, impact is addressed based on two levels: micro and macro, covering, respectively, the individuals and the educational system and society affected by the test.

Practicality

Simply put, practicality of a test according to Farhady, Ja'farpour and Birjandi (2006, p. 159), "refers to the ease of administrations and scoring of a test". Bachman and Palmer (1996) expanded the areas of practicality even to the rationale behind test design. So they define it as "the way in which the test will be implemented, and, to a large degree, whether it will be developed and used at all (p.35)". This definition implies that in deciding on test practicality, the relationship between the required and available resources should be taken into account as practicality is the function of the following schematic representation (Bachman and Palmer 1996, p.36):

$$\text{Practicality} = \frac{\text{Available resources}}{\text{Required resources}}$$

If practicality ≥ 1, the test development and use is practical.

If practicality < 1, the test development and use is not practical.

Therefore, test practicality is judged based on the type and amounts of the required resources for its design, operationalization and administration on

one hand, and the available resources for conducting these three stages on the other.

Further studies have been conducted on test characteristics. According to Henning (1987) " examinations that serve as admissions criteria to university must be highly reliable; whereas a quiz used to determine which children may be rewarded with a longer recess period between classes would be less critical " (p.10). He emphasized that both test should be reliable, but the importance of the decision in the former situation requires greater reliability than in the latter one. For example, Ayers (1979) studied the predictive validity of the National Teacher Examination Test. He approached his purpose by choosing 148 teaching students in university during their first year of education and he started to calculate the correlation between teachers' scores on National Teachers Examination Test (NTET) and the teachers' scores on American College Test (ACT). The results showed the predictive validity of NTET. His study could really assure American's policy makers of the usefulness of NTET and it convinced the authorities in education policy to trust the results of such tests and try to recruit the candidates who earn the highest scores on NTET, in particular.

In another study, Gumaa Siddiek (2010) did a study to explore the features of the Sudan School Certificate English Examination (SSC) from the perspectives of content validity and comprehensiveness. The results of this study showed that SSC English Examination forms are incomprehensive and lack content validity. The author claimed that they are proficiency tests rather than scholastic standardized achievement examination, and, therefore, have negative backwash in language education development in Sudan.

As it is seen, no comprehensive approach similar to that of Bachman and Palmer (1996) has yet been practiced as to language test evaluation. Their model on test 'usefulness' accommodates the six mentioned criteria of test development and seems the most integrative framework for test evaluation as well specially when designing, developing and evaluating high-stake tests like the Iranian State University EFL Entrance Examination Test (UEEET) known as the Concour.

The UEEET is a criterion measure based on which millions of university candidates compete to be admitted for EFL program at university level. This very special and exclusive EFL entrance examination is administered nationwide as an independent instrument to screen the candidates for three subfields of EFL including: English Literature, English Translation, and Teaching English. Known as Concour (from the French; Konkoor, Konkour, and Konkur translated into the Persian), it is claimed as a standardized norm-referenced test used as the means for the candidates to gain admission to higher education in Iran.

Additionally, limited space and resources restricting many talented and enthusiastic applicants seeking access to higher education (Kamyab, 2009) have changed the Iranian State UEEET from a selection into a competition test which forces both teacher and learners to focus too heavily on test preparation at the expense of other activities such as listening and speaking skills. So, some Iranian EFL students frequently complain that their English speaking, listening and even their writing skills are poor, though they seldom complain about knowing grammatical points.

Some parts of such complaints, weaknesses, and dissatisfaction may originate from poor entry criterion measure. Meanwhile, these complains may be attributed to the ignorance of the test taker's characteristics and are believe to be of crucial significance as, in this respect, Henning (1987) stated that "when we faced with the responsibility of having to choose or develop an appropriate test, we should take some matters into consideration, including information as the purpose of the test, the characteristics of the examinees, the accuracy of measurement, the suitability of format and features of the test, the developmental sample, the availability of equivalent

or equated forms, the nature of the scoring and reporting of scores, the cost, the procurement, and the political acceptability of the test"(p.9).

In addition, according to Bachman and Palmer (1996), " test developers must evaluate the usefulness of their tests during all three stages of test development including design, operationalization, and administration. Sometimes, they do it informally, while some other times they use more formal procedures, such as going through checklists and collecting data on test usefulness "(p.133). So, the main problem is then whether the instrument utilized for Iranian State University EFL admission meets terms of a good quality test. In other words, the test Usefulness seems to be a questionable and controversial issue. To this end, the main purpose of this study was to evaluate the Iranian State UEEET in terms of six test qualities including reliability, validity ,impact , interactiveness , authenticity, and practicality as suggested by Bachman and Palmer (1996), from the perspectives of both instructors and students in order to find out its strengths and weaknesses and also help test developers and policy makers to make proper modifications so that they can improve the quality of this nationwide exam in order to achieve a balance among the qualities of test usefulness (Bachman & Palmer, 1996) and set the minimum acceptable levels for each. This problem and purpose were converted into a multi-faceted research question as follows.

Research Question

Main Question: Does Iranian State University EFL Entrance Examination Test enjoy all common characteristics of test usefulness (i.e., reliability, validity, impact, authenticity, interactiveness, and practicality)?, and if so, to what extent?

This seven-fold research question accommodating, in fact, six distinct indices as the criteria of test usefulness plus its extent, were investigated distinctively in the form of seven research hypotheses [though removed here due tospace limitations].

Methodology

Participants

The people who participated in this study were from two different groups including professors and students. The former group included 30 professors, comprising both male and female, from three State Universities in Tehran. All of them were PhD holders in TEFL, English Translation, and English Literature. The latter group consisted of 100 students selected randomly. They were both males and females between the ages of 19 and 21. All participants were EFL major freshmen who had very recently experienced taking the UEEET.

Instrumentation

Two separate but interdependent instruments were employed for the purpose of this study:

1. *A likert-sacle 30-item questionnaire in two versions (i.e., Persian and English) was developed out of the criteria questions suggested by Bachman and Palmer on addressing test usefulness (1996, pp. 150-155) Each item contains four alternatives and the participants were encouraged to choose only one option representing their ideas appropriately. The students received the Persian version , while the professors attempted the English one.*
2. *A semi-structured interview composed of 10 general items abstracting main points and criteria of the Questionnaire. The interview form was represented in Persian in order to ameliorate the process of data collection.*
3. *The 2013 version of the UEEET which had been recently experienced by the students. The Test was attached to the Questionnaire for two purposes: the students to take it once again in order to attempt the questionnaire items very informatively and the professors to take a glance at the Test while attempting the Questionnaire.*

Procedure

Data were collected using both versions of the Questionnaire along with the latest version of the UEEET. Moreover, a semi-structured interview was conducted with volunteer students (N=20) and professors (N=10). The manuscripts of the interview were analyzed in terms of content and major, and common themes were extracted, classified and coded. Mainly frequency analyses along with chi-square test were run for the statistical purposes. The percentages of the participants marking each relevant item to certain test quality were calculated.

Data Analysis and Results

The collected data as to the UEEET usefulness qualities were analyzed both discretely concerning each quality and integratively measuring the usefulness as a whole construct. Meanwhile, all analyses are paired for a clear picture purpose and better comparative judgment. It is worth mentioning that the Questionnaire was composed of four choices (i.e., Very strongly; Strongly; Fairly and Weakly). But given the wordings of the questionnaire items eliciting the ideas of the participants, the two first choices (i.e., Very strongly and Strongly) were the bases of the statistical analyses although the third choice (i.e., fairly) by itself could be a sound support to answer the main and minor research questions.

Reliability

According to Table 1, majority of the teachers (57.7%) and students (55.7%) and totally 56.02% of both groups believe that the UEEET' reliability fluctuates in different administrations.

Table 1

Teachers and Students' Attitude towards Reliability of the UEEET: Frequencies, Percentages and Std. Residuals

			Choices				Total
			Very Strongly	Strongly	Fairly	Weakly	
Group	Teachers	Count	20	66	53	10	149
		% within Group	13.4%	44.3%	35.6%	6.7%	100.0%
		Std. Residual	.9	-.2	.1	-.8	
	Students	Count	49	218	168	44	479
		% within Group	10.2%	45.5%	35.1%	9.2%	100.0%
		Std. Residual	-.5	.1	.0	.4	
Total		Count	69	284	221	54	628
		% within Group	11.0%	45.2%	35.2%	8.6%	100.0%

None of the Std. Residual values are beyond the ranges of +/- 1.96. That is to say, there are not any significant differences between the teachers and students' attitude towards the reliability of the UEEET. In addition, the results of the chi-square (χ^2 (3) = 1.90, P > .05) (i.e. Table 2) indicate that there are not any significant differences between the teachers and students' attitude towards the reliability of the UEEET.

Table 2

Teachers and Students' Attitude towards Reliability of the UEEET: Analysis of Chi-Square

Section		Value	Df	Asymp. Sig. (2-sided)
Reliability	Pearson Chi-Square	1.909[a]	3	.591
a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 12.81.				

Devoted to the data collected through the Interview, Table 3 represents some major themes regarding the test reliability from the teachers' perspectives. According to teachers' ideas, test administration settings in Iran are usually different from one test setting to another with low qualities of the facilities.

Table 3
Teachers' Perceptions of Reliability

Major Themes	(%)
Usually varies from one test administration to another	.50
No item structure variety	.20
Test format is not satisfactory, make students exhausted	.30

As displayed in Table 4, large number of the students believed that the Test setting characteristics are not consistent over various administration of the Test. There are uncomfortable chairs, bad temperature, and sometimes long distance to the place of administration which make candidates exhausted.

Table 4
Students' Perceptions of Reliability

Major Themes	(%)
In consistent conditions with low administration quality	.30
Uncomfortable chairs and noisy	.15
Items are difficult, small in size and too many in numbers	.10
Emphasis on vocabulary and grammar knowledge and reading comprehension	.25
Reponses are confusing especially for vocabulary items	.10
Appropriate administration affect test takers performance	.10

analyses revealed that majority of the teachers and students believed in the Test inconsistency in various administrations. So, the Test is generally evaluated as a less reliable one.

Validity

According to Table 5, majority of the teachers (68.7%) and students (67.04%) believed that the UEEET enjoys acceptable validity.

Table 5
Teachers and Students' Attitude towards the Validity of the UEEET: Frequencies, Percentages and Std. Residuals

			Choices				Total
			Very Strongly	Strongly	Fairly	Weakly	
Group	Teachers	Count	42	61	36	11	150
		% within Group	28.0%	40.7%	24.0%	7.3%	100.0%
		Std. Residual	.4	-.1	-.6	.8	
	Students	Count	125	203	133	26	487
		% within Group	25.7%	41.7%	27.3%	5.3%	100.0%
		Std. Residual	-.2	.1	.3	-.4	
Total		Count	167	264	169	37	637
		% within Group	26.2%	41.4%	26.5%	5.8%	100.0%

None of the Std. Residual values are beyond the ranges of +/- 1.96. That is to say there are not any significant differences between the teachers and students' attitude towards the validity of the UEEET. The results of the chi-square ($\chi 2$ (3) = 1.56, P > .05) (i.e. Table 6) indicate that there are not any significant differences between the teachers and students' attitude towards the validity of the UEEET.

Table 6

Teachers and Students' Attitude towards the Validity of the UEEET: Analysis of Chi-Square

Section		Value	Df	Asymp. Sig. (2-sided)
Reliability	Pearson Chi-Square	1.526[a]	3	.676

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 8.71.

So based on the data presented in Table 7, majority of the teachers argued that special attention should be given to the Test content; indicating that the Test content should be relevant to its purpose and there should not be any contradiction between materials taught in schools and materials presented in the Exam.

Table 7

Teachers' Perceptions of Validity

Major Themes	(%)
Test characteristics consistent to Test purpose	.40
Usually check students perception and understanding	.20
Content and form of the Test affect test takers performance	.20
Items of the test cannot measure test takers ability completely	.20

According to the students' opinions as indicated in Table 8 , test content should be related to materials presented in schools. They mentioned that some of the items of the Test, especially vocabulary section, lack content validity and most students had not even heard some of the vocabulary items presented in the Exam.

Table 8

Students' Perceptions of Validity

Major Themes	(%)
Content is confusing	.10
Content not consistent with educational curriculum	.20
Structure of the Test consistent with its purpose	.15
Check Test takers knowledge of vocabulary, grammar and reading comprehension	.30
Items do not measure test takers production ability	.10
Reduce test takers stress if content consistent to book materials	.15

Therefore, contrary to the index of reliability, both groups believed that the UEEET enjoys acceptable validity characteristics.

Impact

According to Table 9, majority of the teachers (69.08%) and students (66.08%) believe that the UEEET has a strong impact on society, teachers and students.

Table 9

Teachers and Students' Attitude towards the Impact of the UEEET: Frequencies, Percentages and Std. Residuals

			Choices				Total
			Very Strongly	Strongly	Fairly	Weakly	
Group	Teachers	Count	72	154	83	16	325
		% within Group	22.2%	47.4%	25.5%	4.9%	100.0%
		Std. Residual	-.6	1.0	-.6	-.3	
	Students	Count	262	459	300	59	1080
		% within Group	24.3%	42.5%	27.8%	5.5%	100.0%
		Std. Residual	.3	-.6	.3	.2	
Total		Count	334	613	383	75	1405
		% within Group	23.8%	43.6%	27.3%	5.3%	100.0%

Moreover, none of the Std. Residual values are beyond the ranges of +/- 1.96. That is to say there are not any significant differences between the teachers and students' attitude towards the impact of the UEEET. The impact characteristic of the UEEET has three components; impact on

teachers, students and society. Thus, according to Table 10, majority of teachers and students believe that all three sub-sections have strong impacts on teachers, students and society.

Table 10
Components of the Impact: Frequencies and Percentages

Group				Choices				Total
				Very Strongly	Strongly	Fairly	Weakly	
Teachers	Section	Impact on Students	Count	40	87	43	10	180
			% within Section	22.2%	48.3%	23.9%	5.6%	100.0%
		Impact on Teachers	Count	19	40	24	4	87
			% within Section	21.8%	46.0%	27.6%	4.6%	100.0%
		Impact on Society	Count	13	27	16	2	58
			% within Section	22.4%	46.6%	27.6%	3.4%	100.0%
	Total		Count	72	154	83	16	325
			% within Section	22.2%	47.4%	25.5%	4.9%	100.0%
Students	Section	Impact on Students	Count	115	261	178	36	590
			% within Section	19.5%	44.2%	30.2%	6.1%	100.0%
		Impact on Teachers	Count	84	135	72	4	295
			% within Section	28.5%	45.8%	24.4%	1.4%	100.0%
		Impact on Society	Count	63	63	50	19	195
			% within Section	32.3%	32.3%	25.6%	9.7%	100.0%
	Total		Count	262	459	300	59	1080
			% within Section	24.3%	42.5%	27.8%	5.5%	100.0%

The results of the chi-square ($\chi 2$ (3) = 2.42, P > .05) (i.e. Table 11) indicate that there are not any significant differences between the teachers and students' attitude towards the impact of the UEEET on teachers, students and society.

Table 11
Teachers and Students' Attitude towards the Impact of the UEEET: Analysis of Chi-Square

Section		Value	Df	Asymp. Sig. (2-sided)
Impact	Pearson Chi-Square	2.428[a]	3	.488
a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 17.35.				

However, the instructive and informative side of the coin is the nature of the impact. According to Table 12, majority of the teachers referred to the significant consequences of the Test on the students' future life, work, and destiny, such that some candidates who do not pass the Exam, get to be disappointed. In contrast, there were some teachers who believed in positive impact of the Test on the candidates, teachers and society.

Table 12
Teachers' Perceptions of Impact

Major Themes	(%)
Impact can be positive or negative based on candidates purpose	.50
Results are essential and affect test takers future life	.30
Given feedback is not complete	.20

but a little more surprisingly, based on Table 13, majority of the students (85%) hold that the Test consequences are mainly negative and according to them they vary parallel to the candidates' purposes. For those who want to gain experience about the Exam, the impact might be positive but for those who participate in the Exam in order to be admitted to universities and find appropriate jobs, the effects might be negative.

Table 13
Students' Perceptions of Impact

Major Themes	(%)
Negative and cause lots of stress for test takers	.35
Reduce self-confidence	.10
Consequences is positive and identify students, teachers and material strengths or weaknesses	.15
Feedback is not meaningful to test takers	.20
Decisions are not sometimes fair	.10
Vocabulary and grammar items are not consistent with text book materials	.10

Interactiveness

Base on Table 14, majority of the teachers (72.07%) and students (63.5%) or in other words 66.5 % of all participants believe that the UEEET is strongly interactive.

Table 14
Teachers and Students' Attitude towards the Interactiveness of the UEEET: Frequencies, Percentages and Std. Residuals

			Choices				Total
			Very Strongly	Strongly	Fairly	Weakly	
Group	Teachers	Count	43	105	74	14	236
		% within Group	18.2%	44.5%	31.4%	5.9%	100.0%
		Std. Residual	-1.5	.2	1.1	.0	
	Students	Count	192	340	207	47	786
		% within Group	24.4%	43.3%	26.3%	6.0%	100.0%
		Std. Residual	.8	-.1	-.6	.0	
Total		Count	235	445	281	61	1022
		% within Group	23.0%	43.5%	27.5%	6.0%	100.0%

None of the Std. Residual values are beyond the ranges of +/- 1.96. That is to say, there are not any significant differences between the teachers and students' attitude towards the interactiveness of the UEEET.

The interactiveness characteristic of the UEEET has five components; topical know ledge, suitability, language knowledge, cognitive strategies and affective schemata. Majority of teachers and students, as table 15 shows, believe that all five sub-sections are strongly interactive; except for students' attitude towards "language knowledge" section.

Table 15
Components of the Interactiveness: Frequencies and Percentages

Group				Choices				Total
				Very Strongly	Strongly	Fairly	Weakly	
Teachers	Section	Topical Knowledge	Count	4	27	23	4	58
			% within Section	6.9%	46.6%	39.7%	6.9%	100.0%
		Suitability	Count	19	24	15	2	60
			% within Section	31.7%	40.0%	25.0%	3.3%	100.0%
		Language Knowledge	Count	3	12	13	1	29
			% within Section	10.3%	41.4%	44.8%	3.4%	100.0%
		Cognitive Strategies	Count	12	28	14	6	60
			% within Section	20.0%	46.7%	23.3%	10.0%	100.0%
		Affective Schemata	Count	5	14	9	1	29
			% within Section	17.2%	48.3%	31.0%	3.4%	100.0%
	Total		Count	43	105	74	14	236
			% within Section	18.2%	44.5%	31.4%	5.9%	100.0%
Students	Section	Topical Knowledge	Count	31	82	72	13	198
			% within Section	15.7%	41.4%	36.4%	6.6%	100.0%
		Suitability	Count	85	86	25	2	198
			% within Section	42.9%	43.4%	12.6%	1.0%	100.0%
		Language Knowledge	Count	12	24	39	21	96
			% within Section	12.5%	25.0%	40.6%	21.9%	100.0%
		Cognitive Strategies	Count	48	105	38	7	198
			% within Section	24.2%	53.0%	19.2%	3.5%	100.0%
		Affective Schemata	Count	16	43	33	4	96
			% within Section	16.7%	44.8%	34.4%	4.2%	100.0%
	Total		Count	192	340	207	47	786
			% within Section	24.4%	43.3%	26.3%	6.0%	100.0%

The results of the chi-square ($\chi 2$ (3) = 4.76, P > .05) (i.e. Table 16) indicate that there are not any significant differences between the teachers and students' attitude towards the interactiveness of the UEEET.

Table 16
Teachers and Students' Attitude towards the Interactiveness of the UEEET: Analysis of Chi-Square

Section		Value	Df	Asymp. Sig. (2-sided)
Interactive	Pearson Chi-Square	4.76[a]	3	.190
a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 14.09.				

Table 17
Teachers' Perceptions of Interactiveness

Major Themes	(%)
Topical knowledge is usually based on test developers' assumption	.40
Little attention to language knowledge	.20
No opportunity for strategy involvement	.20
No attention to personal characteristics of the test takers	.20

According to Table 18, most students (75%) believed that the UEEET enjoys very negligible level of Interactiveness, but not interactive enough in many of the other areas. That is to say some students referred to the lack of using test items which measure essential language abilities such as speaking, listening and writing.

Table 18
Students' Perceptions of Interactiveness

Major Themes	(%)
Not that much attention to topical knowledge	.30
Little attention to language knowledge	.15
Do not test other types of language knowledge such as listening, speaking, and writing	.25
Limited use of strategies	.10
No care about test takers feelings and their differences	.20

The comparison of the perceptions indicate that all data types are comparable and the participants' ideas are generally compatible , though they differ in some areas of Interactiveness. Compared to the data collected through the questionnaires, those collected through the interview seemed to be more comprehensive.

Authenticity

Based on the results displayed in Table 19, it can be concluded that 72.4% of teachers and 74.4% of students or in other words 67.2 % all participants strongly believe that the UEEET enjoys required authenticity characteristics.

Table 19
Teachers and Students' Attitude towards Authenticity of the UEEET: Frequencies, Percentages and Std. Residuals

			Choices				
			Very Strongly	Strongly	Fairly	Weakly	Total
Group	Teachers	Count	11	10	7	1	29
		% within Group	37.9%	34.5%	24.1%	3.4%	100.0%
		Std. Residual	.6	-.6	-.1	1.6	
	Students	Count	30	44	25	0	99
		% within Group	30.3%	44.4%	25.3%	0.0%	100.0%
		Std. Residual	-.3	.3	.1	-.9	
Total		Count	41	54	32	1	128
		% within Group	32.0%	42.2%	25.0%	0.8%	100.0%

None of the Std. Residual values are beyond the ranges of +/- 1.96. That is to say there are not any significant differences between the teachers and students' attitude towards the required authenticity characteristics of the UEEET. In addition, the results of the chisquare ($\chi 2$ (3) = 4.36, P > .05) (i.e. Table 20) indicate that there are not any significant differences between the teachers and students' attitude towards there required authenticity characteristics of the UEEET.

Table 20
Teachers and Students' Attitude towards the Authenticity of the UEEET: Analysis of Chi-Square

Section		Value	df	Asymp. Sig. (2-sided)
Interactive	Pearson Chi-Square	4.360[a]	3	.225
a. 2 cells (25%) have expected count less than 5. The minimum expected count is .23.				

However, when the areas are studied a bit more discretely, 50% of the teachers held that the test is less authentic. According to teachers' ideas displayed in Table 21, the UEEET is rather consisted to Target Language Use situation because of the correspondence between characteristics of TLU domain and the characteristics of the test items. But in some cases, specially, the test content and expected responses they hold different views.

Table 21
Teachers' Perceptions of Authenticity

Major Themes	(%)
Structure of the Test is not consistent with TLU	.30
Condition of the Test is not consistent with TLU	.20
Expected response is rather desirable	.20
Items usually concerned with students memorization	.30

Meanwhile, according to Table 22, majority of the students referred to the inconsistency of Test condition with TLU situation. They believed that in order to be ascertain about the participants language knowledge, test developers must consider different kinds of language abilities not just few particular types. Then, it could be said that about 70% of the students question the authenticity of the Test.

Table 22
Students' Perceptions of Authenticity

Major Themes	(%)
Condition of the Test not very consistent with TLU	.15
Dot not check all possible knowledge of students	.25
Content are sometimes ambiguous	.10
Multiple-choice is fine	.15
Must be other ways of reposing such as open-ended items and interview section as well	.15
Difficult vocabulary items are used which is not usual in TLU	.20

All data types revealed that the UEEET enjoys the minimum acceptable authenticity level.

Practicality

Since there were no items in the questionnaire addressing the UEEET practicality criterion; therefore, the researcher resorted mainly to the data collected through the interview in order to investigate teachers and students' attitudes towards its practicality.

As to the practicality of the Test, according to Table 23, most teachers (80%) expressed dissatisfaction with regard to the resources and the Test quality. In contrast, designing and writing Test items do not seem a difficult process and also do not cost much for the test writers.

Table 23
Teachers' Perceptions of Practicality

Major Themes	(%)
Resources are not usually available to all the students	.40
Resources are old	.17
Not that much difficult for the test writers	.20
Low qualities of the test papers	.13
Should make some modifications in design of the Test	.10

Variety of ideas regarding the test practicality as shown in Table 24 indicates that nearly all students mentioned that resources of the Test are not usually available to them and do not specify accurately. Roughly, about 50% of the students believed that the Test does not enjoy the main characteristics of practicality.

Table 24
Students' Perceptions of Practicality

Major Themes	(%)
Design of the Test is weak	.12
Resources are not easily available for the students in various regions	.22
Some resources are expensive	.06
For the test writers is not a difficult job	.07
Test items are usually similar to each other yearl	.10
No special recourses for vocabulary and reading comprehension sections	.20
Resources not consistent with items presented in the Test	.14
Low qualities of the test papers	.09

Consequently, majority of the teachers and students believed in low practicality of the Test.

Usefulness

Generally evaluating the Test in terms of the Usefulness Model as a whole, as Table 25 shows, majority of the teachers (65.6%) and students (65.5%) or in other words 65.5% of all

participants strongly believe that the UEEET enjoys usefulness characteristic.

Table 25

Teachers and Students' Attitude towards the Usefulness of the UEEET: Frequencies, Percentages and Std. Residuals

			Choices				Total
			Very Strongly	Strongly	Fairly	Weakly	
Group	Teachers	Count	188	396	253	52	889
		% within Group	21.1%	44.5%	28.5%	5.8%	100.0%
		Std. Residual	-.6	.5	.0	-.1	
	Students	Count	658	1264	833	176	2931
		% within Group	22.4%	43.1%	28.4%	6.0%	100.0%
		Std. Residual	.3	-.3	.0	.1	
Total		Count	846	1660	1086	228	3820
		% within Group	22.1%	43.5%	28.4%	6.0%	100.0%

None of the Std. Residual values are beyond the ranges of +/- 1.96. That is to say there are not any significant differences between the teachers and students' attitude towards the usefulness of the UEEET.

The results of the chi-square (χ^2 (3) = .867, P > .05) (i.e. Table 26) indicate that there are not any significant differences between the teachers and students' attitude towards the usefulness of the UEEET.

Table 26

Teachers and Students' Attitude towards the Usefulness of the UEEET: Analysis of Chi-Square

Section		Value	df	Asymp. Sig. (2-sided)
Reliability	Pearson Chi-Square	.867[a]	3	.833

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 53.06.

Conclusion and Discussion

This study tried to implement Bachman and Palmer's (1996) 'Usefulness Model' in order to evaluate the Iranian State UEEET both quantitatively and qualitatively from the main consumers' (i.e., Teachers and Students) perspectives in a bit to seek their perceptions. The findings supported mainly complementary , though sometimes contradictory results. Both groups evaluated the Test mainly less reliable , while they held positive views towards the validity as sustained by both types of data as well. As to the impact factor (not to be confused with the ISI type Impact Factor!!), both groups held that the Test entails impact, but when delving into the nature and areas of the factor it is revealed not only their perceptions vary but also they are mainly negative indicating that the Test does not bear the desired impact value. Similar evaluation is made and can be claimed as to the Interactiveness and Authenticity qualities since not only the data are not much compatible but also in some areas they are contradictory, which leave the gap for further investigations. Regardless of all these matches and mismatches, the Test has been totally evaluated enjoying the Usefulness features, though there are many issues requiring further peculations

On the complexity of test evaluation and characteristics, Alderson et al. (1995) , it is not possible to achieve a fully reliable test. Thus, the test developers should try to make the UEEET as reliable as possible. They should minimize the effects of those sources of inconsistency that are under the test developers' control, through the test design. However, according to the results, although the UEEET is not much stable and reliable over various administrations, according to Henning (1987),since

validity assumes and entails reliability, it can be concluded that a valid test entails some degree of reliability. As the UEEET had acceptable validity index; therefore, it can be concluded that the UEEET is to some extent is reliable thanks to its validity index.

Validity and especially construct validity, according to Bachman and Palmer (1996) " pertains to the meaningfulness and appropriateness of the interpretations that we make on the basis on test scores " (p.21). Thus, the Test developers should attempt to develop the UEEET items in a way that they can provide the decision makers with appropriate evidence that the Test scores reflect the areas of the language ability they intended to measure. Consequently, as mentioned by Henning (1987), while validity assumes and entails reliability, it is probably more important for the UEEET to be more valid than reliable.

According to the results, although the UEEET has potential impact in the test takers and their characteristics, on teaching and learning activities and on educational system and society, according to Bachman and Palmer (1996), there exists one way to promote the UEEET to have positive impact. They suggested that it would be better to involve the test takers in the design and development of the test, as well as in collecting information from them about their perceptions of the test. In this way, the test tasks are perceived as more authentic and interactive and the test takers probably perform better.

Based on the results, the UEEET relatively involves the test takers' areas of the language knowledge, cognitive strategies, topical knowledge and affective schemata. However, it seems essential for the UEEET to create a balance among the extent and type of the test takers' characteristics in accomplishing the Test. Therefore, unless the UEEET requires a relative involvement of the test takers' areas of language knowledge, strategic competence, metacognitive strategies and affective schemata, it is not possible for the decision makers to make inferences about the language ability of the test takers' performance.

Bachman and Palmer (1996) present two main reasons for considering authenticity as an important test quality. According to them, test authenticity " a) provides a link between test performance and the TLU tasks and domain to which we want to generalize and b) the way test takers' perceive the relative authenticity of test tasks can facilitate their test performance" (p.39). Therefore, it seems essential for the test developers to construct the UEEET in a way that it can provide the correspondence between the characteristics of the test items with the characteristics of TLU domain.

The test developers should do their best to design and develop the UEEET based on the available sources for all the test takers' and, meanwhile, thy should try to allocate the additional sources in order to acquire more appropriate results on the part of the teachers and the test takers. Therefore, according to Bachman and Palmer (1996, p.36) ," a practical test is one whose design, development , and use do not require more resources that are available ". Thus, it seems essential for the test developers to clearly determine the sources that are used in developing and constructing the UEEET for both the teachers and the test takers.

Consequently, all of the six test qualities contribute to the usefulness of the UEEET, and according to Bachman and Palmer (1996), test developers should set an appropriate balance among all of the six test qualities. Thus, in the development of the UEEET, it is essential not to ignore any of these qualities at the expense of the others. The test developers should achieve a balance among all the essential test qualities in order to be able to define the test takers' language ability accurately and appropriately.

References

Alderson , J. C. and Clapham , C. & Wall, D. (1995). *Language Test construction and Evaluation*. Cambridge: Cambridge University Press.

Ayers, J. B. (1979).Concurrent and predictive validity of the national teacher examina-

tions. *Journal of Educational Research*, 73, 25-69.

Bachman, L. F., & Palmer, A. (1996). *Language testing in practice*. Oxford: Oxford University Press.

Brown, J.D., & Hudson, T. (2002).*Criterion-referenced language testing.* Cambridge: Cambridge University Press.

Farhady, H., Ja'farpur, A., &Birjandi, P. (2006). *Testing language skills: from theory to practice*. Tehran: SAMT.

Farhady, H. (2006). *Twenty-five years of living with applied linguistics: collection of articles.* Tehran: Rahnama Press.

Fulcher, G. (2010). *Practical Language Testing.* Great Britain: Hodder Education

Gumma Siddiek, A. (2010).Evaluation of the Sudan school certificate English examination. *English Language Teaching Journal*, 3(2), 37-47. Retrieved from http:// www.ccsenet.org/elt

Henning, G. (1987).*A guide to language testing.* Cambridge Mass: Newbury House.

Kamyab, S. (2009).Junior Colleges in the Islamic Republic of Iran. In Raby, R. L.,&Valeau, E. J. (Eds).*Community college models*. VI (561-568).Springer

Kardan, A. M. (1972). The history of evolution of entrance match of university and higher education institutes in Iran. *Science Education and Psychology quarterly*, 1(2), 70-78. Faculty of Science Education and Psychology: University of Tehran.

Acknowledgements:

Our languages are not eloquent enough to express our sincere appreciation for ***Dr. Mahmood Moradi-Abbasabadi*** from Mazandaran University in Babolsar (Iran) for his insightful and professional comments on the early draft of this paper.

Gholam-Reza Abbasian, an assistant professor of TEFL at Imam Ali University & IAU, has presented at (inter) national conferences, is the author & translator of about 15 books, publisher of scholarly articles and offers psycholinguistics, language testing, & research methods at MA and PhD levels. He has been nominated as top scholar and the most successful teacher for seven consecutive years. He is the internal manager of JOMM, reviewer of Sage, FLA and GJER journals and a member of editorial board of JSSIR.

Hadieh Nasirian, hold an MA in TEFL from IAU (South Tehran Branch). She teaches English in Islamic Azad University (SAMA) and some institutes in Tehran. She is interested in doing research in language testing and assessment, psycholinguistics, and SLA.