

Power Relations among Different Test Parties from the Perspective of Critical Language Assessment

S. Tahmasebi *

Ph.D. Candidate, TEFL

Shiraz Branch, Islamic Azad University

email: ahmasebi_so@yahoo.com

M. Yamini

Assistant Professor, TEFL

Shiraz Branch, Islamic Azad University

email: yaminimortaza@yahoo.com

Abstract

Critical Language Assessment (CLA) argues that language testing is a form of social practice and an agent of cultural, social, educational as well as ideological agendas. As such this article scrutinized the power that different groups of people including, teachers, students and their parents own at the levels of developing, administrating and interpreting Iranian University Entrance Examinations (IUEE). The study revealed that tests could be tools of power that serve the empowered parties' policies and manipulate individual lives. Supported by the findings, the article (1) could be conducive to developing critical thinking among students to see behind the tests, and (2) suggests more democratic testing methods to moderate high-stakes tests consequences so that individual rights get protected.

Keywords: critical language assessment, test parties, power relations, IUEE

1. Introduction

When Messick (1994 and 1996) suggested the argument of “use oriented tests” and explained that tests are prone to affect curriculum, social classes, bureaucracy, politics and knowledge, some doubts were cast on the use of psychometric traits as the only means to judge language tests. The one-dimensionality of psychometric paradigm had to be targeted by taking some steps beyond numbers (Shohamy 2001a); hence, a movement called Critical Language Assessment/Testing (CLA/T) appeared. CLT was indulged in the

Received: 05/16/2012

Accepted: 09/16/2012

* Corresponding author

entanglement of testing perplexes to view tests in social, educational and political contexts (Shohamy, 1998; Kramersch, 1993; Pennycook, 1994) and suggested that the act of language testing is at no time impartial (Shohamy, 1998, 2001a). According to Shohamy (2007), the issue of ‘test use’ which “poses questions about the roles that tests play in education and society” (p 117) is *critical language testing*. CLT does not take tests for granted as it asks questions about the future of individuals taking the test, teachers preparing students to the tests, decisions made on test results, and the materials that appear in the tests.

As one of the crucial issues in language testing, the power of tests is recognized by advocates of CLT (Lynch, 2001; Shohamy, 1997, 2001a and b). As a pioneer in CLT, Shohamy (1997) argues that the huge power of tests is dedicated by some human beings, by those in power, either politically or educationally to control the educational system and to inject specific priorities to the society; hence, she calls such methods of language testing “undemocratic”. On the contrary, in testing situations that test takers’ views are counted, a more democratic environment is expected. Lynch (2001) also argues that the same interconnections between language and social, cultural, and political questions could be represented via considering power relations and consequential validity; such relations are not always explicitly reflected, however.

Bourdieu (1991) perceives an implicit and unwritten contract between test takers and test makers, as the former accepts to be dominated and the latter demands to dominate in order to maintain their present power in the society; hence, the undemocratic use of language tests is implicitly practiced. The manifestations of tests are numerical values which seem so impartial that people rarely question test validities.

Despite the paramount importance of power issues in test development and test administrations, and possibilities that CLA stipulates for publicizing implicit social network, Iranian University Entrance Exam (IUEE) is still among the socially-arcaic events. IUEE event craves critical scrutiny as it is a nationwide and one of the most competitive high-stakes tests known worldwide.

The present research aimed to examine the power relations between different test parties of IUEE following a CLA perspective. In fact, the study questioned the amount of control that three main test parties of students, their parents and teachers own at different phases of IUEE. Although this study targets policy decisions, it neither lines the researcher with particular

political parties, nor marks the study a politically-oriented one, as the first and foremost responsibility that a researcher shoulders is to shed lights on areas which are so interwoven with daily chores that nobody cares about their trustworthiness (Rahimi and Sahragard, 2006; Bloor and Bloor 2007). Furthermore, if we do not intend to safeguard the status quo of the present Iranian language testing, a critical perspective that impugns current testing practices and methods, and offers safeguards for individual test takers might be welcome. (Shohamy (1997) used the term “testing” instead of “assessment”. However, the present study uses the two terms interchangeably).

IUEE is known worldwide as one of the high-stakes tests that entails incurable consequences and momentous implications for people, schools, different organizations and the educational systems. The crux of the matter outcries that individuals’ performances on such a single test, happening only once a year, is the determiner of a range of test takers’ future maneuvers such as their academic life, future job and social class although power relations among test parties are not yet clarified or even searched into. In other words, the process of IUEE includes a wide range of stake holders like students, teachers, and parents on the one hand and those who develop administer and make decisions related to IUEE event on the other hand, so clarifying the amount of control that each party owns at different phases of the event should not be neglected. As one of the social aspects of language tests, power relations are explicitly absent at the expense of numerical values. Researchers, testing experts, and even teachers and students are explicitly negligent in considering test-related social issues, of which power is an example; hence, the power of tests is implicitly imposed.

In the Iranian society, the testing method for entering universities in general, and for evaluating other forms of knowledge in particular, are questioned sporadically by researchers through lenses of wash back effects (e.g., Saif, 2006) and ethical considerations (e.g., Riazi and Razavipour, 2011; Farhady and Hedayati 2009; Farhady, 2006); nonetheless, high-stakes tests are employed to stipulate ministry-set standards and outfit quantitative results with lifelong decisions, as the power of a high-stakes test like IUEE is rarely questioned via academic studies.

Not only are nationwide tests competent to be used as controlling tools handled by political agents or educational elites to derive schools meet the standards and observe their own superficial progress through measurements, but they could be idols not prone to change since no better replacement

could rectify the problematic educational system; furthermore, since such an exam cares about unity and equality in the Iranian multi-ethnic and multicultural society on the face of the matter, it manipulates individuals' social and economic lives with impunity; how much power different stakeholders of IUEE own, how they view the testing method of the event, and whose expectations are realized are problems which require a CLT scrutiny.

1.1 Power and language testing

Pennycook (2001) suggests a highly skeptical perspective for analyzing human freedom, a view different from what Marxists, structuralists, and poststructuralists suggest to the analysis of social and political power, one which cares about the missing link, the individuals. He questions Marxist view that produces relations of power through class relations in which individuals are tools in larger class relations instead of being freed to decide about what they want. This view is similar to Foucault's (1975) as he points to a new modality of power where each test taker obtains his/her individuality since scores demonstrate some features that identify the test taker as a particular "case".

Connecting similar ideas to language learning concerns, Auerbach (1995) asserts decisions related to curriculum development, materials, and language use processes are to be shaped apolitically and based on professional measures, but they impose some ideologies and socioeconomic roles for learners. Spolsky (1997) argues that tests and examinations have inherited power by the moment they were born; therefore, taken the rule that tests are powerful for granted one can ask, how come that tests have developed so powerfully?

Language testing is one of those rare sciences which skew toward empirical procedures, (Farhady 2006, Shohamy 2001; McNamara & Roever, 2006. Shohamy (1997) argues that the power of tests is due to the fact that: (1) test makers own the scores and individuals are surveillant, (2) decision makers refer to test scores to exercise power in educational programs, and (3) scores are assumed to be valid information about test takers' ability since objective measures were implemented to gather them. That objective dresses cater for the power of test is welcome as a rule of thumb. As McIntyre (1984) claims that since decision makers are to adjust means to the ends efficiently, scores are true since they are scientific data seemingly crystalized by numbers. In other words, the test scores are respected since there are no other numbers to rival them (Hanson, 1993); the same

objectivity and appeal to rationality legitimize testing practices (Broadfoot 1996).

1.2 Power and test parties

Rea Dickins (1997) argues that a lot of stakeholders are influenced by test scores ranging from students to teachers, to parents, to administrators, to government agents to funding organizations, to publishers, and to public. Farhady (2006), referring to the event of IUEE classifies all these stakeholders into educational, social and political groups, asserts that each group has a different type of interest in and intention for utilizing tests as a source of power.

The educational group including students, teachers and teacher trainers care for more test quality as they are the most affected group regarding the consequences. Teachers mostly try to concentrate on the subjects that may appear in the test and their own priorities, planning and teaching are shadowed or absent at the expense of would-be tested materials and strategies. It is paradoxical that both students and teachers and sometimes even parents are cognizant that what appears in the IUEE and the practices they render during class hours are not in line with the current methods of language teaching and learning (Barootchi & Keshavarz, 2002; Hamayan 1995; Tsagari, 2004), but reminiscent of traditional methods (e.g. Grammar Translation methods) for language learning and teaching (Mahdavinia & Rahimi, 2011), still they do not hesitate to act according to the test priorities as their intellectual and social power may not go beyond this since they are manipulated to act in certain ways (van Dijk, 2003; Bloor & Bloor, 2007).

The social context embraces diverse groups regarding social class and each group may interpret test scores differently. Farhady (2006) explains that some families might be dogmatic about scores and enforce some changes in the family affairs to provide a better situation for their children to pass IUEE since they know the result affects the future of test takers and the family while some families do not care about the results and the test process goes on without affecting the families' ordinary path.

Political groups own the highest power as they are partial to tests as the most accessible solution, so they make decisions based on scores. They control the educational system, justify their decisions, and impose curricula which entail specific materials and textbooks through tests (McNamara, 1996; McNamara & Roever, 2006). For example, in some countries like Canada and Australia, governments use language ability tests as a criterion

to control immigration, while in Iran the whole entrance exam, of which language section is just a portion, controls students' desires and persuades them that what happens is a fact predestined by students themselves, not the educational system.

The relationship among the members of each party is another dilemma. Test makers are usually organizations and agencies that move together toward predetermined objectives while test takers are individuals who compete against each other and follow their own concerns (Shohamy, 1998). It is hypothesized that this policy is injected by the in-power party to create a competitive atmosphere and assures the applicants that the only way to success is outscoring others as there is no cut score for the test.

2. Research Questions

In examining the unequal power that different parties dealing with tests own and considering the problems that were mentioned, the following questions were formed;

1. Is it possible to highlight the power relations among different stakeholders of IUEE? How?
2. Are test takers and test developers, who are two important parties in the social context of the tests, equally powered?

3. Context of the Study

Iranian students spend 6 years in elementary and pre-elementary schools, from age 6 or 7 to the age of 12 or 13. Then they move to guidance school and study for three more years. English is included in their instructional courses from the first year of guidance school.

After the four-year high school studies, which also include one year for pre-university study, Iranian students can enter universities. Entry to preferred academic majors and specific universities, particularly state universities, is ultra-competitive, controlled by IUEE which is set by Sanjesh (evaluation) Organization. The format, time of administration and even registration deadline are fixed across the whole country. In fact, insufficient capacity and fund for free academic studies have escalated this predicament.

As one of the areas tested, English has been taught about 7 years at schools. The course books, teaching timetables and even the style guidelines are set by Iranian Ministry of Education which are in line with the government policy that assumes since students are equal in their rights, they

should benefit from equal methods of teaching, instructional materials, and testing system.

Like many other societies, the Iranian cultural system skews toward meritocracy and egalitarian values which are mirrored in the arduous competitive examination systems. Students who work hard and are diligent can outscore the others (Akiyama, 2003). So the underlying construct might be hard work not communicative proficiency. The grueling test should be so difficult to correctly select diligent and talented students. Those who do not get a preferable rank at IUEE cannot study in their desired major; they have to continue studying, mostly memorizing the tests and increasing their speed in answering multiple-choice items, to hopefully act better next year. They struggle to enter universities in order to get the job they dream of. In Iran, those who do not have academic degrees have difficulties in getting a desired job; moreover, there are growing wage gaps between university graduates and non-graduates, let alone the fields of study and the type of university students have studied in.

4. Framework of The Study

Shohamy (2001a) mentioned fifteen principles characterizing critical language testing which are also summarized by Lynch (2001: 363) into four groups. Although all the principles are interconnected and in line with critical movements, it is not possible to implement all the principles in a single study; hence, just principles 5, 8, 11, 13, and 15, since connected to power issues, are applied in the present study.

5. CLT asks questions about which and whose agendas tests serve.

8. CLT examines the influence and involvement of the range of stakeholders in a testing context.

11. CLT considers the meaning of test scores within this interpretive framework, allowing for the possibility of discussion and negotiation across multiple interpretations

13. CLT challenges the primacy of the 'test' as assessment instrument and considers multiple procedures for interpreting the knowledge of individuals.

15. CLT challenges the knowledge that test is based upon and advocates a democratic representation of the multiple groups of the society.

5. Methodology

5.1 Participants

To consider the power issue in language testing of IUEE, the study has invited representatives of three social groups who receive the tests and are affected by the test results including students, teachers, and parents. To have a sample similar to the population, the students were selected from four different regions with different L1s; they were Turkish, Kurdish Arabic and Persian L1 speakers, their age mean was 17.4 and they presented both sexes equally. Teachers are divided into two groups: One group taught English to students whose L1 was Persian, and the other group taught English to students whose L1 was one of the minority languages such as Turkish, Arabic, and Kurdish. Students, their parents and teachers are selected from non-private schools. The ethnic groups were selected via a stratified random sampling method (Dornyei, 2003, 2007), randomly selected from specific L1 categories. About 60 students from each ethnic group as well as their parents and 60 teachers (15 representatives for each L1) participated in the study. The total number of participants was 360. All the mentioned groups could roughly be called non-test makers party, as they neither make the test nor have roles in making decisions on different phases of IUEE.

5.2 Instruments

The instrument used in this study was a questionnaire developed to yield answers to three types of factual, behavioral and attitudinal questions. It originally had an open-ended format to permit greater freedom of expression. A semantic differential scale was used where a characteristic statement preceded five boxes (1=have no idea; 2=never/strongly disagree; 3=seldom/ slightly agree; 4= frequently/often agree 5=always/strongly agree). Since the questionnaire was developed to specifically gather information about the amount of control that each test party would have in developing and administrating IUEE, feasible indices for validity and reliability were not available, hence following Dornyei (2003, 2007) some measures were taken to guarantee the internal consistency, i.e., the homogeneity of items.

First, using multi-item scales, a cluster of differently worded items that focused on the same target was written. Then, to maximize the stable component, which the items share and to reduce the extraneous influences unique to individual items, more than one item was used to address identified content areas, i.e. different wordings were used to present the

same concept. The wording was simple and item root rarely exceeded eight words and usually covered half of the line. Both negatively and positively worded items were used which in turn reduced the acquiescence bias.

To check the internal consistency, Cronbach's Alpha Coefficient was used second to piloting the items to a group of respondents sharing most similarities with the target sample.

5.3 Pilot study

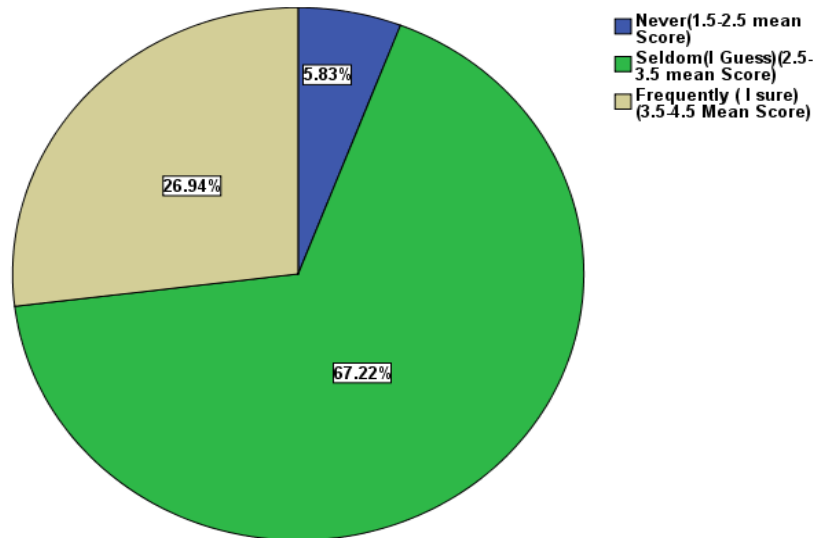
To reduce the dependability of questionnaires on actual wording (Dornyei 2003), the questionnaire received field test on 80 samples as akin as possible to people for whom the instrument was created.

The result of Cronbach's Alpha analysis showed that the internal consistency of all items was above 90%, hence reliable. However, the analysis required us to delete some items to increase internal consistency so that an overall consistency of .915 became evident, and the items were reliable to be used as an instrument for gathering information.

5.4 Data analysis

To analyze the views of different participants with regard to their ethnicity and position, first one-sample Kolmogorov-Smirnov test was used to check if the distribution of data was normal. The overall distribution of scores was normal and the data were qualified for further analyses. The Pie Chart 1 depicts the distribution of the data.

Figure 1. The distribution of responses regarding the frequency or agreement



To compare the amount of power among different groups of participants, a one-way ANOVA was used. Table 1 shows that there existed a meaningful difference among the groups regarding the power as $p < 0.0001$.

Table 1. ANOVA results on power among groups

Power control	Sum of Squares	Df	Mean Square	F	Sig.
Between Groups	7.139	6	1.190	6.095	.000
Within Groups	68.908	353	.195		
Total	76.047	359			

Post hoc results of Bonferroni test further indicated that the difference between some groups of participants is significant, although with different magnitude. The results are presented in Table 2.

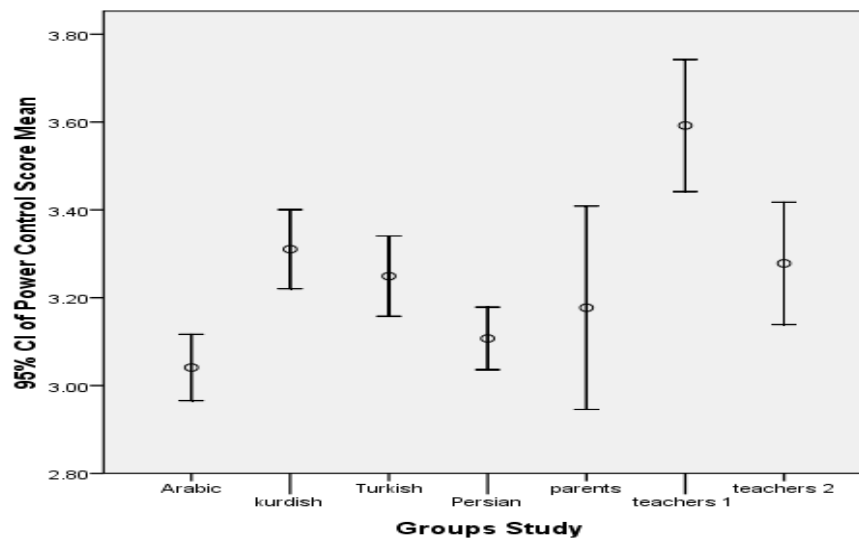
Table 3. Post hoc results of the differences between the groups

(I) group	(J) group	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Arabic	Kurdish	-.26935	.08085	.020	-.5168	-.0219
	Turkish	-.20803	.07825	.172	-.4475	.0314
	Persian	-.06610	.07473	1.000	-.2948	.1626
	Parents	-.13612	.11730	1.000	-.4951	.2229
	Teachers 1	-.55093	.10627	.000	-.8762	-.2257
	Teachers 2	-.23709	.10108	.411	-.5464	.0722
kurdish	Arabic	.26935	.08085	.020	.0219	.5168
	Turkish	.06132	.07542	1.000	-.1695	.2921
	Persian	.20325	.07176	.103	-.0163	.4228
	parents	.13322	.11543	1.000	-.2200	.4865
	Teachers 1	-.28158	.10420	.152	-.6005	.0373
	Teachers 2	.03226	.09890	1.000	-.2704	.3349
Turkish	Arabic	.20803	.07825	.172	-.0314	.4475
	Kurdish	-.06132	.07542	1.000	-.2921	.1695
	Persian	.14193	.06883	.838	-.0687	.3526
	Parents	.07191	.11363	1.000	-.2758	.4196
	Teachers 1	-.34290	.10221	.018	-.6557	-.0301
	Teachers 2	-.02906	.09679	1.000	-.3253	.2672
Persian	Arabic	.06610	.07473	1.000	-.1626	.2948
	Kurdish	-.20325	.07176	.103	-.4228	.0163
	Turkish	-.14193	.06883	.838	-.3526	.0687
	Parents	-.07002	.11123	1.000	-.4104	.2704
	Teachers 1	-.48483	.09954	.000	-.7894	-.1802
	Teachers 2	-.17099	.09397	1.000	-.4586	.1166
Parents	Arabic	.13612	.11730	1.000	-.2229	.4951
	Kurdish	-.13322	.11543	1.000	-.4865	.2200
	Turkish	-.07191	.11363	1.000	-.4196	.2758
	Persian	.07002	.11123	1.000	-.2704	.4104
	Teachers 1	-.41481	.13447	.046	-.8263	-.0033
	Teachers 2	-.10097	.13040	1.000	-.5000	.2981
Teachers 1	Arabic	.55093	.10627	.000	.2257	.8762
	Kurdish	.28158	.10420	.152	-.0373	.6005
	Turkish	.34290	.10221	.018	.0301	.6557
	Persian	.48483	.09954	.000	.1802	.7894
	Parents	.41481	.13447	.046	.0033	.8263
	Teachers 2	.31384	.12058	.202	-.0552	.6828
Teachers 2	Arabic	.23709	.10108	.411	-.0722	.5464
	Kurdish	-.03226	.09890	1.000	-.3349	.2704
	Turkish	.02906	.09679	1.000	-.2672	.3253
	Persian	.17099	.09397	1.000	-.1166	.4586
	Parents	.10097	.13040	1.000	-.2981	.5000
	Teachers 1	-.31384	.12058	.202	-.6828	.0552

According to Table 2, the difference between the following two groups is significant:

Arabic versus Kurdish ($p < 0.020$); For Arabic versus Teachers 1, (teachers who teach English to L1 Persian), $p < 0.000$; Turkish versus Teachers 1 $p < 0.018$; Persian versus Teachers 1 $p < 0.000$; Parents versus Teachers 1 $p < 0.046$. The difference between the remaining pairs of the groups is not significant; for example, regarding the Kurdish speakers, there is just a significant difference between this group versus Teachers 1, and Turkish, while in other cases there is no significant difference between Kurdish L1s and the other groups as $p < 1$ between Parents and Kurdish group. These results may further indicate that all groups of participants do not share equal amount of control toward different phases of IUEE development and administration. The difference between the means of all groups is also depicted in the following Figure.

Figure 1. Interactive Graph displaying the difference between groups



Descriptive statistics are presented in Table 3. Since the Confidence Interval is 95%, it could be concluded that predictions are up to 95% error free. As it is shown, almost all participants have a mean about 0.3; it indicates that the participants have taken the safe side, the middle ground. It may further show that participants in most cases are not well informed of the time of the test development, test writers, sources of the questions and the system of testing, and they just guess the event.

Table3. Descriptive statistics

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean	
					Lower Bound	Upper Bound
Arabic	56	3.0411	.34354	.04591	2.9491	3.1331
Kurdish	64	3.3104	.43772	.05472	3.2011	3.4198
Turkish	74	3.2491	.47846	.05562	3.1382	3.3599
Persian	93	3.1072	.41743	.04329	3.0212	3.1931
Parents	19	3.1772	.61344	.14073	2.8815	3.4729
Teachers 1	25	3.5920	.45687	.09137	3.4034	3.7806
Teachers 2	29	3.2782	.45594	.08467	3.1047	3.4516
Total	360	3.2133	.46025	.02426	3.1656	3.2610

6. Discussion

To highlight the amount of power that different stakeholders own, three groups of students (test takers from four different L1s), their parents, and teachers (from two different contexts; one group taught English to Persian L1s, and the other taught English to non-Persian L1s), were asked to fill out the questionnaires. Taking an eye-bird point of view, all these groups are test receivers, neither test developers nor test administrators. In fact, it was not possible to access the first major party named test makers and/or test developers as well as test administrators, since they are either anonymous or impossible to access to due to safety measures.

Question No 1: *Is it possible to highlight the power relations among different stakeholders of IUEE? How?*

Yes. The findings of the study suggest that CLT can be an appropriate method to highlight power relations among test parties as it raises questions about test parties' control over tests. First, it is possible to examine the influence and involvement of a range of stakeholders in a testing context (Principle 8). The questionnaire asked several questions to check the amount of involvement of test takers, teachers and parents as three groups of test parties who receive the impacts of IUEE test. The results show that more than 30% of the participants had no information about test makers (the mean of the answers to five questions on the participants' information about test developers was estimated); 34% had no information about the time that the test is made (the mean of answers to two questions).

Next, CLT challenges the dominant psychometric traditions and considers 'interpretive' approaches to assessment that allow different

meanings and interpretations rather than a single abstract entity (Principle 11). The fact that about 40% of the participants believe the present testing method of IUEE is modern, and 22% percent agree with the present multiple-choice format may further reveal the position different groups of test receivers are in. It may go well with what Shohamy (2001b, 2007) suggests that some practices may evolve into ideologies as test receivers do not challenge the ruling method of evaluation, or they are manipulated to do so (Bloor & Bloor 2007); they are convinced that the present method of testing is the best to impartially differentiate among them. The participants have also agreed (more than 35%) that questions should not match learning styles and individual differences. In other words, they have accepted the view that if all participants receive one test, their rights are respected. Through CLT, it is possible to raise doubts about such contradictions among test parties views; these results (views regarding the format of the test and equity of test) show that the method has been practiced so intricately that participants rarely doubt its policies. Moreover, as there are different parties in test development, the public have to trust scores which seem impartial (Farhady, 2007). In fact, a range of measures should determine the success or the failure, not a single performance. CLT makes it possible to critically question the dominant testing methods, and check the control that each test party exerts over the high-stakes test of IUEE. Furthermore, CLT challenges the knowledge upon which the test is based and advocates a democratic representation of the multiple groups of the society; 60% of participants agreed with involving students, and 63% agreed with involving school teachers in test development. These views contradict with what participants have suggested about testing method, too (30% agreement with present testing method of IUEE), since the percentage of their agreement with democratic practices in testing is twice their views regarding the main stream method of IUEE other arguments are possible. They are willing to have hands in test preparation and moderating the consequences of tests; however, as they have accepted IUEE as an unchallenged destiny, they have no further ideas regarding other possible alternatives. Democratic assessments make it possible not to get succumb to the ruling main stream system, and protect individuals against unethical practices in language testing (Shohamy, 2001b).

Still more, CLT challenges the primacy of the 'test' as assessment instrument and considers multiple procedures for interpreting the knowledge of individuals (Principle 13). The results of the study showed that 60% of

the participants prefer to be tested by their own schools and teachers not a centralized testing method. Since the knowledge that tests measure is relative not completely true, Shohamy (1998) suggests that students' achievement be considered as art, as long as the results could be relative, interpersonal and interpretive.

Finally, CLT challenges the knowledge that test is based upon and advocates a democratic representation of the multiple groups of the society. A single test may manipulate every layer of the educational system with its dictatorial power to fantasize the society with its charisma (Principle 15). The results show that although the six groups of participants who have answered the questionnaire have indicated that they are not aware of different processes of IUEE (including test makers, time of test development, sources of reading comprehension texts, as well as the expectations that are met), they prefer the present testing method. The justification for taking such positions might be that since they are not raised in a democratic society, they cannot imagine democratic principles in a testing situation; they easily take the safe side and seek their asylums in scores and multiple choice formats which seem impartial to them. What has clarified such positions is the CLT perspective that this study has taken.

To put it in a nutshell, since the mentioned data highlighted the amount of control that different test parties hold in different processes of IUEE, it may suggest using CLT in examining use-oriented tests and provide researchers the spotlight and courage they require to highlight different aspects of language tests, of which power issue is an example.

Question No 2: *Are test takers and test developers, who are two important parties in the social context of the tests, equally powered?*

The results of the analysis suggest “no” to this question. There was a significant difference between the answers that each group of participants provided; moreover, the following points rejected the null hypothesis:

1. The results suggest that test takers do not share equal or even close opinions about the method of language testing in IUEE, and due to the diversity of the participants' social and educational background some rights are expectedly ignored. Although almost 36% of the participants assert that the mainstream method of IUEE is never modern (Table 4), 30% of Persian L1 speakers and 31% of Arabic L1 speakers strongly agree that it is modern. On the other hand, regarding the format (multiple-choice versus open ended), 28% of the participants (21% strongly, and 26% often) agreed with an MC format for the test (Q.22)

while 43% agreed (21% strongly, 22% often) that both formats are adequate to be used in IUEE. Now, how is it possible to test them via a single equated test to supposedly care for rights and diminish the differences? Can we consider the rights of different people who have to take the same test for different purposes if they answer to the same test at a particular time? The results of this study suggest that the method of testing should invite all participants' preferences and views.

2. Auenbach (1995) argues that regarding the dynamics and inequality of power that show up even in curriculum development and evaluation, one should know about people who form the questions, the questions themselves, and the evaluation system. The mean of responses to questions related to test developers suggested that 33% of participants don't know who develops the tests for IUEE, and more than 20% think (always and frequently) that test developers live in big cities like Tehran. It is said that test items are based on the materials which are included in high school textbooks, while 45% of the participants have no information about the sources of questions. Such results also suggest that test parties of IUEE are not equally powered, as test takers, parents and even teachers are not well-informed of even test makers or the sources of questions; such blurry answers to questions related to test developers and test sources provide nothing but the participants' unawareness which in turn suggest lack of power in test development and decisions about the sources. Similarly, Coleman, Starfield, and Hagan (2003) found that the staff of IELTS were not aware of the meaning of test scores, although the students of their study were well aware; however, the results of this study call for further attention to invite more stakeholders, mainly test takers, in order to have a greater democratization in IUEE development and administration.
3. The finding of the study also showed that 48% of the participants agree that present testing method of IUEE is always or frequently preferred over other methods. It could be inferred that the respondents of this study agree that nation-wide tests designed to measure the students' success or failure could improve the quality of education since students have to study more. Such ideas go with Shepard (1991)'s conclusions that tests control and dictate what to be included in the curricular and learnt by the students. It is ironic that while not all schools may have access to materials based on which tests are made and there are mismatches between what teachers are supposed to teach, what they actually teach,

and how teaching abilities of teachers and language abilities of students are evaluated in different areas, public opinions skew toward the virginity of the IUEE testing method. On the contrary, to highlight the problems of such superficial satisfaction, Critical Thinking suggests that we should avoid oversimplification, consider other interpretations, and tolerate ambiguity (Wade, 1995); some research approaches that can entice changes in the society which favor public interests should be acknowledged, second to informing people of the test power that might be manipulated.

4. Another set of questions is asked about those whose expectations are mostly realized through IUEE. Regarding Q No. 6 which asks if questions are made to meet parents expectations, 52.8% never agree, while in answering Q No. 91 more than 52 % of the participants agree (28.9% always and 23.9% almost) that test results meet talented students' expectations. Furthermore, 62% of the participants agree (29.2% always and 33.3% almost) that test takers who know how to answer MC format items get a good score in IUEE. These results go with what Spolsky (2007) suggests that "tests are lotteries biased in favor of the more talented" (p. 213), since the expectations of less privileged parties are ignored or superficially met as the results of this study show that 51% of participants believe that the expectations of all participants are Never or Seldom realized through the current method of testing (Q.5).
5. When the data of the present study suggest that the content, time, method of testing, and value of all sections in IUEE are not decided upon by students, teachers and their parents, the event is controlled by other parties. SO in addition to Iranian Ministry of Education, students are just informed about the irreversible decisions. The rules regarding the development and the administration of the test have slightly changed during the history of IUEE event. Those (teachers) who make the test are always assigned by the SO. Participants have no hand in assigning them, and have less chance in knowing them; the statistics suggested for No.1 above confirm this point. Students just have access to the test and the numerical results, not human beings who have implemented such mechanical devices. These facts increase the gap between stakeholders regarding their control over IUEE. Students and teachers never decide about the consequences of test, as students take the tests and then are informed numerically about the results, teachers and parents are also affected as they have to act according to the test and prepare the students

for the tests, and share the consequences with students. It is argued that the awareness of the power of tests motivates the authorities to inject tests as instruments to control, manipulate and impose specific knowledge to students, and teachers in order to regulate principles and educational systems according to certain agendas and educational ideologies (Broadfoot, 1996; Shohamy, 2001).

7. IUEE may just meet the expectations of SO or the Iranian Ministry of Education. After each exam, some top members of these two organizations appear in public and congratulate the properly-administered event of IUEE; as they appear and appreciate participants and administrators for what they have done in the day of exam, it is inferred that they are the owner and the decision makers of IUEE who control the event. Farhady (1998) argues that if decision makers were test makers or vice versa, the decisions would have been made more cautiously since administrators, bureaucrats and politicians do not feel testing problems, and assume scores as true indicators of abilities; hence, decisions deserve test takers. It could be inferred from these statements and similar related arguments (Pennycook 2001, Shohamy, 1998, 2001) that decision making problems could be alleviated if the test takers and test makers get closer during different stages of test making and administration. When it is possible that language testers question the authoritarian views leveled against test scores, the educational atmosphere is a more desirable one.

In general, the summarized points may suggest that test constructs may be performed as aspects of social, cultural and political priorities and may implicitly lead to power conflicts among stakeholders; moreover, the English language section of IUEE is seemingly devised to measure language abilities while the results of performances are used to make other decisions irrelevant of language use. Hence, to understand the nature of such a test better, we need to consider the multi-dimensional power as well as social consequences by-produced in the existing test constructs. Since the concerns of IUEE could not be tackled through thoroughly conducted psychometric research, as it is the outcome of different social inter-relations, long term consequences, psychological impressions, and power struggles may be highlighted through CLT practices.

The present study was designed in the domain of Critical Language Testing to detect the place of test takers and language teachers in language test development, and challenge the psychometric traditions. That's why the findings might be significant to the field of applied linguistics including

testing and evaluation, teaching, and methodology, as well as curriculum and materials development.

First, basic modifications in evaluation measures and testing practices might be implications of the study. There has been a reconsideration and slight reconstruction of the program by agents in the Iranian Ministry of Education to make evaluations more effective and efficient. The unequal power of test parties revealed through the findings of the present study may help them design and develop their new evaluation programs in line with the democratic testing.

Second, the findings of the study are significant for the modification of assessment and teaching techniques. It seems that students have accepted some aspects of IUEE as an unquestionable fact since they have not been informed of perceiving the depth of issues. According to Wade (1995), Critical Thinking (CT) involves asking questions, defining a problem, examining evidence, analyzing assumptions and biases, avoiding emotional reasoning, avoiding oversimplification, considering other interpretations, and tolerating ambiguity. The analogy of these items to language teaching and testing enhances the teachers' awareness of the entanglement of language testing. Learners could be assisted in improving their critical thinking through the actions that teachers do, e.g. through some consciousness raising.

Third, regarding curriculum and materials development the study suggests a learner-based approach which also entails cognitively supported methods, as the roles of learners are ignored in different processes of IUEE. That is, the syllabi would be a posteriori and retrospective one, open to further negotiation during teaching and testing phases, not a prescribed a priori one dictated by elites, and adorned and implemented by teachers and students (Weir, 2005). The present curriculum has been resistant to change for about forty years. The primary goal of the study was to promulgate critical thinking issues in language evaluation and policies which could enhance self-actualization and learner autonomy. All these cognitively supported and socially valued aspects are almost absent at different phases of language planning and implementation. The results of the study may convince curriculum writers to include materials which indulge the overall thinking and creativity of learners instead of assigning them the role of knowledge consumers (Shohamy, 2000; Pennycook, 2001), who have to memorize some materials which will appear in the EUEE.

The problem highlighted in this study also targets researchers who have not been as outspoken in questioning language test quality as they were morally missioned to and professionally responsible for. The language testing history in Iran reveals that there has been failures to challenge the validity aspect of language section of IUEE, reluctance to consider the dynamic nature of language ability, unwillingness to welcome new trends, ignorance of alternative ways to testing, hesitation to at least modify the testing process, which is resilient in nature and resistant to change.

7. Conclusion

Advocates of critical movements argue that understanding the ways that social practice flows through language, and how issues of power, often obscure in language research and educational practice, are realized in learning practices are all crucially important in developing critical language pedagogy. A particular look at the term critical questions dominant testing practices (e.g. Shohamy, 2001a, 2007; Lynch 2001) and suggests ways in which democratic principles can be applied to assessment practices in multi-ethnic societies. While dominant groups may give lip service to equality principles, the de facto situation in many societies is that tests can serve as tools to maintain and perpetuate the dominant knowledge and policies of in-power groups.

This study argued that although language tests integrate two poles, namely test makers and test takers, and marginalized groups like teachers and parents are also affected, the agenda of language testing reflects that of the in-power party, or the elite who are test makers and test administrators; this fact is almost implicit and imposed delicately. The explicitly-affected parties including teachers, students, and even their families are ignored or slightly invited to the scene (Lynch, 2001, Shohamy, 1998, 2001a, b; Pennycook, 2001). The results of this study revealed that the agenda of IUEE are almost unknown to test takers, their families and the teachers; the event of IUEE is created by more politically-and/or-educationally empowered parties. Considering the fact that these three major groups of stakeholders in language testing have no control over the content, the time of administration and even consequences imply that power is not equally shared among all test parties.

In recent years, there has been growing focus on language testing regarding the importance of consulting with stakeholders and taking their views more seriously (Rea-Dickins, 1997). Therefore, studies like this may

not only remind us the forgotten place of individuals in language testing, but also based on the principles of critical pedagogy they may give language learners a new perspective to understand themselves and their surrounding world as they are, not as they are suggested to be. CLT encourages students not to respect the charisma of tests and test developers/administrators, but to question the agendas, policies, and purposes behind them; to perceive the rights that are ignored; to find the traces that change practices into ideologies (Shohamy, 2001).

Despite power inequalities among test parties and along with concerns about decisions made based on language tests, test developers have to prepare tests and test takers have to get prepared for taking them. Moreover, as it is said, tests are here to stay (Shohamy, 2007) and never could studies like this challenge the legitimate rights that tests benefit from, but the whole discussion ends in shedding some lights on the issue and devising other policies that individual rights are observed and authoritarian positions are replaced by alternative forms of assessment which skew toward more democratic decisions.

References

- Akiyama, T. (2003). Assessing speaking in Japanese junior high schools: Issues for the senior high school entrance examinations. *JALT Journal*, 7(2).p. 2–11. Retrieved Jan. 12, 2012 from jalt.org/test/PDF/Akiyama1.pdf
- Auerbach, E. (1995). The politics of the ESL classroom: Issues of power in pedagogical choices. In J. Tollefson (Ed.). *Power and Inequality in Language Education* (pp. 933). Cambridge: CUP.
- Barootchi, N., & Keshavarz, M.H. (2002). Assessment of achievement through portfolios and teacher-made tests. *Educational Research*, 44, 279-288.
- Bloor, M. and Bloor, Th. (2007) *The Practice of Critical Discourse Analysis. An Introduction*. London: Hodder Arnold.
- Bourdieu, P. (1991) *Language and symbolic power*. Oxford: polity Press.
- Broadfoot, P. (1996) *Education, assessment and society*. Buckingham: Open University Press
- Coleman, D., Starfield, S., & Hagan, A. (2003). Stakeholder perceptions of the IELTS test in three countries. *IELTS Research Reports*, 5, 159-235. Canberra: IELTS Australia. Retrieved April, 12, 2011 from www.ielts.org/pdf/Vol5_Report4.pdf

- Dornyei, Z. (2003). *Questionnaires in second Language Research*. Mahwah NJ: Lawrence Erlbaum.
- Dornyei, Z. (2007). *Research methods in applied linguistics: quantitative, qualitative and mixed methodologies*. Oxford: Oxford University Press.
- Hanson, F. A. (1993). *Testing testing: social consequences of the examined life*. Berkeley: University of California Press.
- Hamayan, E. V. (1995). Approaches to alternative assessment. *Annual Review of Applied Linguistics* 15, 212-226.
- Kramsch, C. (1993). *Context and culture in language teaching*. Oxford: Oxford University Press.
- Farhady, H. (1998). A critical review of the English section of the BA and MA University Entrance Examination. *Twenty-five years of living with applied linguistics: Collection of articles*. Farhady, H. (2006). Iran, Tehran. Rahnama press.
- Farhady, H. (2006). *Twenty-five years of living with applied linguistics: Collection of articles*. Iran, Tehran. Rahnama Press.
- Farhady, H., & Hedayati, H. (2009). Language assessment policy in Iran. *Annual Review of Applied Linguistics*, 29, 132–141.
- Foucault, M. (1975). *Discipline and punish: The birth of the prison*. New York: Vintage.
- Fowler, R. (1991). *Language in the News: Discourse and Ideology in the Press*. London: Routledge.
- Lynch, B. K. (2001). Rethinking assessment from a critical perspective. *Language Testing*, 18, 351–372. Retrieved Jan.12, 2012, from www.ling.lancs.ac.uk/groups/crile/docs/crile58tsagari.pdf
- Mahdavinia, M. Rahimi, A. (2011). An evaluative case study: Reflection-focused assessment performance. *On line Journal Language Testing in Asia* 1(3) 7-33
- McNamara, T. (1996). *Measuring second language performance*. London: Longman.
- McNamara, T., & Roever, K. (2006). *Language testing: The social dimension*. Oxford: Blackwell.
- McIntyre, A. (1984). *After virtue: A study in moral theory*. Notre Dame, Indiana, University of Notre Dame Press.
- Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessment, *Educational Researcher*, 23, 13–23.

- Messick, S. (1996). Validity and wash back in language testing, *Language Testing*, 13(4), 241–57.
- Pennycook, A. (1994). *The cultural politics of English as an international language*. London: Longman
- Pennycook, A. (1999). Introduction: Critical approaches to TESOL. *TESOL Quarterly*, 33(3), 329-48.
- Pennycook, A. (2001). *Critical applied Linguistics: A critical introduction*. London: Routledge.
- Rahimi & Sahragard. (2006,). A critical discourse analysis of euphemization and derogation in e-mails on the late Pope. *The Linguistics Journal*, 2.
- Rea-Dickins, P. (1997). So why we need relationship with stakeholders in language testing? A view from UK. *Language Testing*. 14 (3), 304-314.
- Riazi, A.M. & Razavipour, K. (2011). Agency of EFL teachers under the negative backwash effect of centralized tests. *International Journal of Language Studies (IJLS)*, Vol. 5(2), 122-142. Retrieved Jan. 12, 2012, from www.ijls.net/volumes/volume5issue2/riazi1.pdf
- Saif, S. (2006). Aiming for positive washback: a case study of international teaching assistants. *Language Testing*, 23 (1), 1-34.
- Shepard, L. (1991). Psychometricians' beliefs about learning. *Educational Researcher*, 20 (7), 2-9.
- Shohamy, E. (1997). Testing methods, testing consequences: Are they ethical? Are they fair? *Language Testing*, 14(3), 340–349.
- Shohamy, E. (1998) Critical language testing and beyond, *Studies in Educational Evaluation*, 24(4), 331–345.
- Shohamy, E. (2001a) Democratic assessment as an alternative. *Language Testing*, 18(4), 373–392. Retrieved Jan., 12, 2012 from www.people.iup.edu/.../Democratic%20Assessment%20as%20an%20...
- Shohamy, E. (2001b). *The power of tests: a critical perspective on the uses of language tests*. Harlow: Longman.
- Shohamy, E. (2007) Language tests as language policy tools. *Assessment in Education* 14, (1), 117–130. Retrieved March 20, 2012, from <https://cmap.helsinki.fi/servlet/SBReadResourceServlet?rid...1TD>
- Spolsky, B. (1997). The ethics of gatekeeping tests: What have we learned in a hundred years? *LanguageTesting*, 14, 242–247.
- Spolsky, B. (2007). Towards a theory of language policy. *Working Papers in Educational Linguistics*, 22(1),1-14.

- Tsagari, D. (2004). Is there life beyond language testing? An introduction to alternative language assessment. CRILE Working Papers, 58. (Online) Retrieved March, 10, 2011, from <http://www.ling.lancs.ac.uk/groups/crile/docs/crile58tsagari.pdf>.
- Van Dijk, T. A. (2003). The discourse-knowledge interface. In G. Weiss, & R. Wodak (Eds.). *Critical Discourse Analysis: Theory and Interdisciplinarity*. (pp. 85-109). Houndsmills, UK: Palgrave-MacMillan.
- Wade, C. (1995). Using writing to develop and assess critical thinking. *Teaching of Psychology*, 22, 24-28.
- Weir, C. J. (2005). *Language testing and validation: An evidenced-based approach*. Basingstoke, ENG: Palgrave Macmillan.