Teachers' Interpretations and Power in a High-Stakes Test: A CLA Perspective

Soheila Tahmasbi (Corresponding author)
Shiraz Branch, Islamic Azad University, Iran
E-mail: Tahmasebi_so@yahoo.com

Mortaza Yamini
Shiraz Branch, Islamic Azad University, Iran
E-mail: yaminimortaza@yahoo.com

Received: September 9, 2012 Accepted: October 20, 2012 Published: November 25, 2012
doi:10.5430/elr.v1n2p53 URL: http://dx.doi.org/10.5430/elr.v1n2p53

Abstract

As one of the high-stakes tests known world-wide, Iranian University Entrance Exam (IUEE) influences the lives of a range of people including EFL teachers. This study, conducted within the framework of Critical Language Assessment (CLA), examined the power of EFL teachers in IUEE and their interpretations of language scores therein. The results of factor analysis demonstrated that high-school teachers have no power in IUEE development and administration processes. Teachers also interpreted tests not as indicators of language ability or knowledge but as test-taking skills. Finally, the study reserves the rights for teachers to be involved in the process of IUEE and suggests interpretive approaches to assessment that allow for different meanings and interpretations rather than a single score.

Keywords: CLT, IUEE, CLA, Test power, Test interpretation, Critical testing

1. Introduction

The moment one thinks of classrooms, schools, and learning, the term teachers inevitably comes to one's mind. Scores and exams are no exceptions in reminding us of our school teachers, as the type of the test and the justifications for taking the exam are connected with the particular teacher who developed the test. For teacher-made tests, the meaning of scores could be interpreted in place. Failing means one has not learnt the materials, while passing denotes mastery of the subject. Although such correspondence between scores and meaning is not always as straightforward as it seems, at least the test, the test-maker, and the test-taker are all present to interact and provide immediate feedback. However, when one talks about high-stakes tests, where the impacts are irreversible, school teachers' power and rights in test development and test score interpretations are paradoxically ignored.

Issues like power of tests and test score interpretations assert that language assessment is first and foremost a social and political activity (Spolsky, 1995; Shohamy, 2001, 2007; McNamara & Roever, 2006) which also integrates language use (Messick, 1996). In examining testing methods which measure language use and reflect the contest between social values of test parties, psychometric traditions which prioritize objectivity are not proper means. A perspective which goes beyond numbers, and introduces use-oriented tests is required (Madaus, 1988; Messick, 1996, 1998).

Shohamy (1998, 2001a, 2001b) defines the issue of 'test use' which casts doubts on the roles that tests play in educational and social contexts. The new outlook raises questions about which and whose agendas tests serve. It examines tests in terms of their measurement and assessment of knowledge versus their definition and dictation of knowledge. Some CLA followers call it a paradigm shift which introduces a testing culture (Wolf, Bixby, Glenn, & Gardner. 1991; Birenbaum, 1996; Pennycook, 2001; Lynch, 2001), since a handful of methods provide evidence for high-stakes decisions not a single test. More importantly, CLA moves toward a democratization in language testing since students and teachers as two immediate test parties are no longer ignored.

Among the principles of CLA (Lynch, 2001; Shohmy, 1998) is the possibility to examine the influence and involvement of a range of factors in a testing context, to consider power relations among test parties, and to take

interpretive approaches to assessment that allow for different meanings and interpretations rather than a single abstract truth.

The high-stakes nationwide test of Iranian Universities Entrance Exam (IUEE) so powerfully influences many spheres of social and academic life that the event could be a relevant subject to current developments and challenges in language testing and assessment. The possibilities that CLA stipulates for publicizing implicit social network of testing events and allowing for negotiations across multiple interpretations, could be extended to IUEE as it is still among socially non-researched events. IUEE is one of the most competitive tests known worldwide wherein each year more than one million applicants compete to obtain an acceptable rank to enter state universities. The format, time of administration and even registration deadline are stable across the whole country, and controlled by an organization called Sanjesh (evaluation) Organization (SO). In fact, insufficient capacity and fund for free academic studies has escalated this predicament.

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 washback effects (e.g., Saif, 2006) and ethical considerations (e.g., Riazi & Razavipour, 2011; Farhady & Hedayati, 2009; Farhady, 2006). Nonetheless, high-stakes tests are ubiquitously employed to outfit irreversible quantitative results to lifelong decisions with impunity. Furthermore, teachers' role in IUEE process is almost ignored as the exam does not seek teachers' views in developing the test, choosing testing method and making decisions. The test does not follow what is taught or should be taught at schools, but it conversely determines what and how of teaching.

Although it is not possible to shed lights on different aspects of language teachers' professional and ethical concerns regarding IUEE in a single study, examining teachers' power in test development and administration, their interpretations of test scores and the interaction between these two issues might have some implications for test parties including teachers. Therefore, the study explored answers to the following research questions based on what language teachers suggested through a questionnaire specifically developed for the purpose of the present study.

1) How much power do language teachers have in test development, test content and test method of IUEE?

2) How do language teachers interpret language scores of IUEE?

3) Is there any relationship between the interpretations that language teachers assign to test scores and the amount of power that they hold in test development and administration?

Following CLA, the study aimed to challenge the primacy of the 'test' as assessment instrument to the high-stakes test of IUEE, and introduce multiple procedures for interpreting the knowledge of individuals. The study further aimed to suggest teachers' involvement in test development and unveil language teachers' ignored roles in high-stakes tests. Through CLA, the study aimed to unfold construct-irrelevant issues which may accompany test scores and lead teachers to interpret test scores differently from what SO does.

2. The Framework

Shohamy (2001) mentioned fifteen principles characterizing critical language testing which are also summarized by Lynch (2001: 363) into four groups. As this study aimed to find out about language teachers' power in IUEE processes and their interpretations of the language scores when the results are announced, principles 8, 11 and 12 are taken into account.

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

Principle 11. CLA challenges the dominant psychometric traditions and considers 'interpretive' approaches to assessment that allow for different meanings and interpretations rather than a single abstract truth.

Principle 12. CLA considers the meaning of test scores within this interpretive framework, allowing for the possibility of discussion and negotiation across multiple interpretations (Lynch, 2001: 363).

3. Language Teachers and Evaluation Issues

While teachers' roles in educational programs are always welcomed, it is imperative to extend their effective roles beyond the present mainstreams of testing, especially in centralized educational systems. Teachers are immediately connected to test-takers and the impacts of IUEE are not limited to end of the course scores, but to lifelong repercussions that determine the university one should attend, the major one is presumably qualified for, and the jobs one can obtain. Regarding national high-stakes testing programs, the bulk of the studies related to teaching and testing come to the conclusion that teachers focus higher attention on previously tested parts (e.g., McMillan, Myran, & Workman 1999; Abrams, Pedulla, & Madaus, 2003; Ballard & Bates, 2008). Some studies end in suggestions for

teachers to develop proper tests (Rudner & Schafer, 2002), or point to teachers' poor knowledge base in assessment (Razavipour, Riazi & Rashidi, 2011).

Studies which have acknowledged teachers' power in language testing and their role in the interpretation of the results are limited in number. As one of the first attempts, Rea Dickins (1997) argued that a lot of stakeholders are influenced by test scores ranging from students to teachers, to parents, to administrators, to government agents, to publishers, and to the public. He found that these parties are different regarding their awareness of test development and test score interpretation. Similarly, Farhady (2006), referring to the event of IUEE, classified all test stakeholders into educational, social and political groups and asserted that each group has a different type of interest in and intention for utilizing tests as a source of power.

Teachers mostly try to concentrate on the subjects that may appear in the IUEE and their own priorities, planning and teaching are overshadowed at the expense of what students will be tested on (Farhady, 2006; Tsagari, 2004; Shohamy, 2007b). Nonetheless, teachers are cognizant that what appears in the IUEE and the practices they render during class hours are not in line with current methods of language teaching and learning (Barootchi & Keshavarz, 2002; Hamayan, 1995; Wolf et al., 1991; Tsagari, 2004). Yet, they do not hesitate to act according to test methods priorities. As long as IUEE recourses to test results to make one of the most important decisions in almost all individuals' lives, similar practices continue.

As one of the educational parties, teachers might get succumbed to the impacts of social, psychological and environmental factors such as school requirements, society's expectations, local policies, mandated curriculum, the practices of peers, workload, and the availability of resources, (e.g., Breen, Caine, & Coltheart, 2001; Khonamiri & Salimi, 2010). Such practices may implicitly serve the expectations of the system since teachers are required by parents and even local agents to do so. Riazi and Razavipour (2011) argue that in centralized systems "teacher's agency is attenuated or even denied in favor of conformity to the dominant structure". They conclude that Iranian system of education "represents a reductionist approach toward education" which assigns a "lowered and disempowered role for teachers" (p. 123).

Connecting these hallmarks to studies which considered the power of tests (e.g., Madaus, 1988; Spratt, 2005; Shohamy, 2001a, 2007a), it can be concluded that testing determines what to be taught, not the official curriculum. IUEE organizers sometimes invite teachers to take part in the evaluation process by asking them to suggest some test items. However, as Danielson and McGreal, (2000) revealed, in such rare cases the evaluative criteria are already fixed, and teachers have no opportunity to contribute to the development of the criteria.

Coffman (1983) argues that what affects teachers' reaction to tests is the extent to which policy decisions by public agencies depend on test results. Similar to Shohamy (2001a), he continues that there should be "guards against assigning so much weight to the results of the test" (p.13), which may even determine teachers' professional practices. It seems that such test-driven practices have manipulated cultural, social and even ideological preferences in the Iranian society as even teachers who are among the literate agents of the society may unconsciously comply with ruling expectations.

4. Method

4.1 Participants

From 80 English language teachers who were invited, 74 agreed to participate in the study. These 74 teachers were teaching at senior high schools in different educational districts of Khuzestan Province, Iran. Forty teachers were teaching at private language schools and/or institutes besides teaching at state-run public schools. Thirty teachers spoke both Persian and Arabic, and the rest spoke Persian as their native language. All the participants had English as their foreign language and were full time tenure employees of the Ministry of Education. The mean age was 34, and the mean teaching experience was about 12 years. Ten teachers had MA in TEFL and the rest held BA degrees in three majors of English Language Teaching, Translation or Literature. As almost all English teachers in major cities of Khuzestan participated in this study, such demographic information may further show the diversity of Iranian English language teachers at high school level in terms of their degree, teaching experience, age, and educational district. Thanks to the support and cooperation given by the head of Khuzestan Language Teachers' Association, it was possible to have views of almost all English language teachers as the data were collected in the English teachers' formal gatherings where all the teachers were present.

4.2 Instruments

The instrument used in this study was a questionnaire specifically developed to yield answers to two types of factual and attitudinal questions. First, 10 open-ended items were developed (Dörnyei, 2003), the answers to which

provided a range of data to be categorized for the ultimate questionnaire. To avoid writing statements which might not be easy and to the point, a semantic differential scale was used where a characteristic statement preceded 5 boxes. Respondents would be asked to indicate their level of agreement by marking one of the boxes between two bipolar adjectives (absolutely agree to completely disagree) on the extremes. Although semantic differential scales have more limitations than likert scales, they were preferred due to some merits. They are adoptable to concepts like power and interpretations used in this study. Another bonus of such scales refers to the fact that they require little reading, and hence less time is required to answer them. Since the questionnaire included 34 items, it seems rational to assume that the quality of the responses was not threatened by boredom and time pressure.

Since the questionnaire was developed to specifically gather information about language teachers' amount of power in test development and administration and the meaning that the results of the test convey to them, following Dörnyei (2003, 2007), some measures were taken to guarantee the internal consistency i.e., the homogeneity of items. For example, using multi-item scales, a cluster of differently worded items that focused on the same target were written to average the idiosyncratic interpretation of an item. That is, individual items were not overloaded and if one item were missed, the damage would not be serious. To maximize the stable component, which the items shared and to reduce the extraneous influences unique to individual items, more than one item addressed certain content areas, i.e., different wordings presented the same concept. The wording was simple and item root hardly exceeded the other half of the line, and rarely exceeded eight words.

Overall, 40 items were developed which integrated two issues of power and interpretation. However, after the pilot study, 6 items had to be removed to increase internal consistency. Of the remaining 34 items, 19 targeted power issues and examined teachers' involvement in test development, their information about test-makers, their views regarding the present method of testing and their awareness of materials used as exam sources. On the other hand, possible interpretations like viewing scores as representatives of language ability, test-taking abilities, hard work, and instruction methods formed the second part of the questionnaire which included 15 items. The items of the two sections were randomly mixed. To validate the questionnaire, seven university teachers were asked to provide their views regarding Content Validity Index (CVI) of the items which led to some slight changes in the wording of the questionnaire items. Furthermore, the validity of the questionnaire was confirmed by factor analysis, and to check the internal consistency, after piloting the items to a group of respondents sharing most similarities with the target sample, Cronbach's Alpha Coefficient was used (Table 1).

4.3 Results

The participants' answers to the 34-item questionnaire were subjected to statistical analysis procedures using SPSS 16. The results of the factor analysis identified 3 factors as the most important ones related to power. These factors were interpreted by authors as: 1- test development, 2- test content and 3- test method. Table 2 reveals these factors and their related descriptive statistics

The extent to which teachers as one of the parties have information about or control over the methods of test administration, content of the test and testing method (Shohamy, 2001a) may clarify the amount of power that this party holds in IUEE. For example, test development factor included eight items which examined participants' awareness of test developers' place of living, time of test development, and whether university teachers developed the test. Based on the mean, it can be concluded that the teachers neither have a role in test development, nor any power to influence it. As for the content and method, they seem to have a neutral attitude, they neither agree nor disagree.

Regarding the second question of the study which was on teachers' interpretation of the scores, the results of the analysis also identified 3 factors among the 15 items which were related to: 1) knowledge 2) test taking ability, and 3) test score. Table 3 reveals these factors and related descriptive statistics.

Based on the means, we can conclude that the teachers have a neutral attitude towards the point that scores of language section of IUEE are indicators of knowledge or test scores reflect a person's power. However, they agree that test scores indicate test-taking ability. That is, the majority of the teachers believe that test results reflect the test-taking ability (test-wiseness) of the candidates. Test-wiseness refers to participants' capability to use test-taking strategies to select correct answers in multiple-choice tests without having the ability or using the skill being tested (Allan, 1992).

To further check the relationship between the amounts of power that language teachers held in test development, test content, and testing method on the one hand and the interpretations that they give to language test scores on the other

hand, Pearson correlation method was used. The results revealed a low correlation (r = .486, p< .0001). Table 4 gives more information.

5. Discussion

Connected to 'critical discourse analysis' agenda to uncover and draw attention to power relations and to highlight the views and stands of the less powerful, critical language testing introduces an innovative perspective on language testing. CLT framework offered the possibility to expose two aspects of the IUEE to critical scrutiny.

5.1 Research question 1: How much power do language teachers have in test development, test content and test method of IUEE?

The findings of the present study revealed that the 74 surveyed teachers neither have a role in test development, nor any power to influence language tests. As for the content and method, they seem to have a neutral attitude, they neither agree nor disagree. These findings may argue for teachers' involvement in IUEE processes. Freedman (1993) concluded that to work collaboratively with schools and communities, it is necessary to involve teachers, administrators, students and parents in assessment practices. Fetterman et al. (1996) also promoted the notion of 'empowerment evaluation' which involves collaboration of all stakeholders in evaluation programs to foster learning and assessment. All these arguments may question the one-dimensionality of IUEE where objectivity accounts for life-long decisions. The findings can suggest the possibility of integrating alternative assessments (Ellis, 2003) which lead to democratic assessment (Shohamy, 2001a, 2007a) in IUEE. In democratic assessment, according to Shohamy (2001a), local groups like test-takers, students, teachers and schools share power by collecting their own assessments, using multiple assessment procedures like portfolios, self-assessment, projects, observations and tests. Such forms of assessment can provide rich evidence for making inferences and decisions. Following democratic assessment, the professional tester serves mostly as a facilitator who assists other parties in the strategies of collecting the information and in its interpretation (Shohamy, 2007a).

5.2 Research question 2: How do language teachers interpret language scores of IUEE?

Based on the results, teachers have a neutral attitude toward scores of language section of IUEE as indicators of knowledge; however, they agree that scores indicate test-taking ability. Such views may raise doubts about the deterministic decisions that are made on the basis of the results of IUEE. Based on the results, it could be implied that IUEE is too narrow in interpreting scores since participants of IUEE have to acquire other skills, in addition to developing their English proficiency skills in order to achieve a satisfactory score (Allen, 1992). Moreover, as Bachman (1990) asserts, if test method affects test scores, this "facet" might be a potential source of error in measurements, which creates two problems: the error is not recognized, and the inferences to be made are ambiguous. It seems that in the case of IUEE the error to be minimized is already known, as the results of this study suggest; however, test scores are referred to as deterministic evidence to make decisions about candidates' academic majors without considering all parties and factors.

Reporting test scores apart from acknowledging background factors (e.g., student mother tongue; public expectation) can force individuals to draw invalid inferences from test scores (Gipps, 1994). Kane (1992) found that scores on a reading comprehension test could be interpreted as a "measure of skill at answering passage related questions" (p. 533), or as an indicator of verbal ability. Expectedly, as long as different parties play roles in language testing events, different purposes are served due to the authoritative power of tests. These purposes are not explicitly known to all and may assign new meanings to test scores.

Critically supported movements can mobilize teachers and test-takers to get aware of hidden agendas and covert power relation in testing (Pennycook, 2001; Rahimi & Sahragard, 2006). The findings of this study imply that before an inference regarding an examinee's ability test developers and users need clarify what a test aims to measure (e.g., various language components, performance on tasks) and whether it actually measures what it intends to measure (Bachman, 1990). Scores might be representative of test-taking abilities, because examinees' performance may be affected by factors other than their language ability (McNamara, 1996).

5.3 Research question 3: Is there any relationship between the interpretations that language teachers assign to test scores and the amount of power that they hold in test development and administration?

The results revealed a low correlation (r = .486) between teachers' power and their interpretations of test scores. Such outcomes may further imply that teachers' disempowerment in different processes of IUEE is connected to their interpretations of language scores. Sometimes the high-stakes nature of tests perpetuates teachers to behave in certain ways to increase their students' test scores and such behaviors determine their interpretation of how best a high score might be achieved (Shohamy, 2007b). In order to use language tests effectively, Shohamy (2007b) asserts

that the interaction between test constructors and teachers should be boosted and teachers should be trained in assessment measures. Teachers' present status and views about IUEE should be seriously taken into consideration. As Al-Amri (2010) states, teachers who make a serious commitment to using criteria, training in the use of scales, and identifying and reflecting on their previous experiences, power, control and subjective decisions can plausibly anticipate more valid, reliable and practical assessments.

6. Implications

CLA as a critical approach to language testing, like critical traditions in applied linguistics, emphasizes transparency of testing events through making as much information available as possible about the test, explaining the reasons for test administration, its construct, scoring method and content, to those related to the tests (Weideman, 2007). According to Shohamy (2001a), all stakeholders such as policy makers, test-writers, students, parents and teachers should be involved in test development and administration. The effects of tests on teachers, students, institutions, and society are accordingly considered one type of validity evidence as many researchers (Bachman, 2005; Cronbach, 1988; Kane, 1992; McNamara & Roever, 2006; Shohamy, 2001; Pan, 2009) have also stressed the importance of justifying test use and investigating its consequences.

According to one of the tenets of CLA, since the knowledge of testers is incomplete, additional knowledge sources are needed to provide more valid explanation and interpretations of knowledge, therefore, based on the results, teachers as one of test parties who can immediately influence test-makers should be empowered in IUEE processes. Moreover, following CLA tenets, it is possible to guard against one-shot tests for assessing knowledge and use multiple procedures in order to have valid bases for interpreting the knowledge of individuals and groups. These highlights are within reach if power of tests in making high-stakes decisions are known to all test parties, including teachers.

CLA principles may expand teachers', students' and researchers' understanding of the nature of critical language assessment and assign them responsibilities to go beyond quantitative chains in language tests. According to the results, teachers who are one of the socially and educationally literate parties do not believe that scores represent language ability; nonetheless, the results of IUEE are bases for the most important decision for a IUEE participant; therefore, it is possible to challenge the imposed meaning that a single score injects to the society, and to widen the interpretive perspective by involving teachers' interpretations as well. Teachers as one of the marginalized parties of IUEE process, are considered in this study to possibly revitalize the important role of teacher agency (Riazi & Razivipour, 2011) in language assessment since in the democratic paradigm of testing teachers and test-takers are not ignored.

References

Abrams, L.M., Pedulla, J.J., & Madaus, G.F. (2003). Views from the classroom: Teachers' of statewide testing programs. *Theory Into Practice*, 42(1), 18-29. http://dx.doi.org/10.1207/s15430421tip4201_4

Al-Amri, M. (2010). Direct spoken English testing: Is still a real challenge to be worth bothering about. *English Language Teaching, 3(1).* Retrieved May12, 2012 from www.ccsenet.org/journal/index.php/elt/article/download/.../4339

Allan, A. (1992). Development and validation of a scale to measure test-wiseness in EFL/ESL reading test-takers. *Language Testing*, *9*(2), 101-122. http://dx.doi.org/10.1177/026553229200900201

Bachman, L. F. (1990). *Fundamental considerations in language testing.* Oxford: Oxford University Press.

Bachman, L. (2005). Building and supporting a case for test use. *Language Assessment Querterly, 2*, 1-34. http://dx.doi.org/10.1207/s15434311laq0201_1

Ballard, K., & Bates, A. (2008). Making a connection between student achievement, teacher accountability, and quality classroom instruction. *The Qualitative Report*, *13*(4), 560-580. Retrieved Sept.12, 2011from http://www.nova.edu/ssss/QR/QR13-4/ballard.pdf

Barootchi, N., & Keshavarz, M.H. (2002). Assessment of achievement through portfolios and teacher-made tests. *Educational Research*, 44, 279-288. http://dx.doi.org/10.1080/00131880210135313

Birenbaum, M. (1996). Assessment 2000: Towards a pluralistic approach to assessment. In M. Birenbaum, & F.J.R.C. Dochy (Eds.), *Alternatives in assessment of achievements, learning processes, and prior knowledge*, (pp. 3–29). Dordrecht, Netherlands: Kluwer. http://dx.doi.org/10.1007/978-94-011-0657-3_1

Breen, N. Caine, D., & Coltheart, M. (2001). Mirrored-self misidentification: Two cases of focal onset dementia. *Neurocase, 7*, 239-254. http://dx.doi.org/10.1093/neucas/7.3.239

Coffman, W. E. (1983). Testing in the schools: A historical perspective. CSE Report 53. Retrieved July 10, 2011 from WE Coffman - annual invitational conference of the Center for the …, 1983 - cse.ucla.edu

Cronbach, L.J. (1988). Five perspectives on validity argument. In H. Wainer (Ed.), *Test Validity* (pp. 3-17), Hillsdale, NJ: Erlbaum.

Danielson, C. & McGreal, T. L. (2000). *Teacher evaluation to enhance professional learning*. Princetone,NJ: Educational Testing Service.

Dörnyei, Z. (2003). *Questionnaires in second Language Research*. Mahwah NJ: Lawrence Erl- baum.

Dörnyei, Z. (2007). *Research methods in applied linguistics: quantitative, qualitative andmixed methodologies.* Oxford: Oxford University Press.

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford: Oxford University 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. http://dx.doi.org/10.1017/S0267190509090114

Fetterman, D., Kaftarian, S. and Wandersman, A. (1996). *Empowerment evaluation*. Thousand Oaks, CA: Sage.

Foucault, M. (1977). *Discipline and Punish. The Birth of the Prison*. London: Penguin Books.

Freedman W.S. (1993). Linking large-scale testing and classroom portfolio assessments of student writing. *Educational Assessment 1*, 27–52. http://dx.doi.org/10.1207/s15326977ea0101_3

Gipps, C. V. (1994). *Beyond Testing: towards a theory of educational assessment*. London: Falmer Press.

Hamayan, E. V. (1995). Approaches to Alternative Assessment. *Annual Review of Applied Linguistics,* (15), 212-226. http://dx.doi.org/10.1017/S0267190500002695

Kane, M. (1992). An argument-based approach to validity. *Psychological Bulletin, 112*(3), 527-535. http://dx.doi.org/10.1037/0033-2909.112.3.527

Khonamiri,F., & Salimi, M. (2010). The interplay between EFL high school teachers' beliefs and their instructional practices regarding reading strategies. *Novitas-ROYAL (Research on Youth and Language),* (1), 96-107 Retrieved Sept.12, 2011 from www.novitasroyal.org/Vol_4_1/khonamri_salimi.pdf

Lynch, B. (2001). Rethinking assessment from a critical perspective. *Language Testing, 18*(4), 351-372. http://dx.doi.org/10.1177/026553220101800403

Madaus, G. F. (1988) The influence of Testing on the Curriculum. In L. N. Tanner (ed), *Critical Issues in Curriculum: 87th Yearbook for the National Society for the Study of Education,* (pp. 83-121). Chicago: University of Chicago Press.

McNamara, T. (1990). Item response theory and the validation of an ESP test for health professionals. *Language Testing, 7,* 52-76. http://dx.doi.org/10.1177/026553229000700105

McNamara, T. (1996). *Measuring second language performance*. London: Longman.

McNamara, T., & Roever, K. (2006). *Language testing: The social dimension*. Oxford: Blackwell.

McMillan, J., Myran, S., & Workman, D. (1999). The impact of mandated statewide testing on teachers' classroom assessment and instructional practices. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada. Retrieved May, 2012 from www.eric.ed.gov/ERICWebPortal/recordDetail?accno=ED431041

Messick, S. (1996) Validity and washback in language testing, *Language Testing, 13*(4), 241–57. http://dx.doi.org/10.1177/026553229601300302

Messick, S. (1998). Validity, in: R. L. Linn (Ed.) *Educational measurement,* (pp. 13-103). The American Council on Education and the National Council on Measurement in Education.

Pan, Ch. (2009). Insights in Language Testing: An interview with Jessica Wu. *JALT Testing & Evaluation. 13*(2), 9–14. Retrieved May 12, 2012 from jalt.org/test/wu_pan.htm

Pennycook, A. (2001). *Critical applied Linguistics: A critical introduction*. London: Rutledge.

Rahimi, A. & Sahragard, R. (2006). A critical discourse analysis of euphemization and derogation in e-mails on the late Pope. *The Linguistics Journal, (1), 29-86.*

Razavipour, K., Riazi, A. & Rashidi N. (2011). On the interaction of test wash-back and teacher assessment literacy: The case of Iranian EFL secondary school teachers. *English Language Teaching, 4*(1), 156-161.

Rea-Dickins, P. (1997). So why we need relationship with stakeholders in language testing? A view from UK. *Language Testing, 14* (3), 304-314. http://dx.doi.org/10.1177/026553229701400307

Riazi, A.M. & Razavipour, K. (2011). Agency of EFL teachers under the negative backwash effect of centralized tests. *International Journal of Language Studies (IJLS), 5*(2), 122-142. Retrieved Jan.10, 2012 from www.ijls.net/volumes/volume5issue2/riazi1.pdf

Rudner, L., & W. Schafer (2002). *What Teachers Need to Know About Assessment*. Washington, DC: National Education Association. Retrieved July10, 2011 from *LM* Rudner… - 2002 - math.nie.edu.sg

Saif, S. (2006). Aiming for positive washback: A case study of international teaching assistants. *Language Testing*, *23* (1), 1-34. http://dx.doi.org/10.1191/0265532206lt322oa

Shohamy, E. (1998). Critical language testing and beyond, *Studies in Educational Evaluation 24,* 331–45. http://dx.doi.org/10.1016/S0191-491X(98)00020-0

Shohamy, E. (2001a). Democratic assessment as an alternative. *Language Testing*, *18*(4), 373-391.

Shohamy, E. (2001b). *The power of tests: A critical perspective on the uses of language tests*. London: Pearson.

Shohamy, E. (2007a). Language tests as language policy tools. *Assessment in Education, 14*(1), 117-130. http://dx.doi.org/10.1080/09695940701272948

Shohamy, E. (2007b). Washback from language tests on teaching, learning and policy: evidence from diverse settings. *Assessment in Education, 14*(1), 1–7. http://dx.doi.org/10.1080/09695940701272682

Spolsky, B. (1995). *Measured Words: The development of objective language testing.* Oxford: OUP.

Spratt, M. (2005). Washback and the classroom: The implications for teaching and learning of studies of washback from exams. *Language Teaching Research, 9*(1), 5-29. http://dx.doi.org/10.1191/1362168805lr152oa

Tsagari, D. (2004). Is there life beyond language testing? An introduction to alternative language assessment. *CRILE Working Papers, 58*. Retrieved March10, 2011 from http://www.ling.lancs.ac.uk/groups/crile/docs/crile58tsagari.pdf

Weideman, (2007). Overlapping and divergent agendas: writing and applied linguistics research. *Living through languages: An African tribute to Rene Dirven. Stellenbosch: African Sun Media*. 147-163. Retrieved May12, 2012 from www.freewebs.com/weideman/tuk01521OverlapDivergAgendas.pdf

Wolf, D. J., Bixby, J. Glenn, J., & Gardner, H. (1991). To use their minds well: Investigating new forms of student assessment. *Review of Research in Education, 17*, 31-74.

Table 1. Reliability of the questionnaire

Cronbach's Alpha	N of Items
.824	34

Table 2. Descriptive statistics on teachers' powers in IUEE

Factors	*N*	Items	*Min.*	*Max.*	*Mean*	*SD*
Development	74	15,20,40,56,61,71,73,85	1.12	4.50	2.9324	.88175
Content	74	75,76,77,83,91	2.20	4.40	3.4622	.54690
Testing Method	74	24,30,32	1.67	5.00	3.4685	.82007

Table 3. Descriptive statistics on teachers' interpretations of scores

Factors	*N*	Items	*Min.*	*Max.*	Mean	*SD*
Knowledge	74	8,12,13, 24,34	2.20	4.60	3.41	.48465
Test-taking ability	74	1,2,11	2.67	5.00	4.19	.58911
Test score	74	3,5,18,19,22,23,33	2.14	4.86	3.61	.52231

Table 4. The correlation between teachers' interpretations and power

		Teachers' interpretations
Teachers' power	Pearson Correlation	.486**
	Sig. (2-tailed)	.000
	N	74