



Critical language assessment: students' voices at the heart of educational system

Zahra Javidanmehr and Nasser Rashidi

Department of Foreign Languages and Linguistics, College of Humanities and Social Sciences, Shiraz University, Shiraz, Iran.

ARTICLE INFO

Article history:

Received: 5 June 2011;

Received in revised form:

17 July 2011;

Accepted: 27 July 2011;

Keywords

Critical Language Assessment,
Critical pedagogy,
Peer assessment,
Consequential validity.

ABSTRACT

This study attempted to bring Critical Language Testing principles into practice by means of distributing the power, traditionally hold by teachers as the only source of knowledge, in a more unprejudiced way. The study took advantages of three features of fifteen CLT principles proposed by Shohamy (2001). Based on the first feature, which is encouraging an active, critical response from test-takers; learners could develop an assessment scale including five components which was used in the subsequent steps in their peer assessment. In line with the second principle, which is admitting to the limited knowledge of any tester and the need for multiple sources of knowledge, learners moved toward the leading edge taking the control of assessment process to some extent. Peer assessment as one criterion of democratic assessment was applied. The third principle exploited in the study was considering 'interpretive' approaches to assessment that allow for different meanings and interpretations rather than a single absolute truth. Students' scores were reported by both quantitative and interpretive modes with some suggestive sentences. In this way learners' voices are validated. After that and in line with "consequential validity" as one component of alternative assessment, it went through completing the process by bringing testees' problematic areas into the teaching syllabus in a systematic way. With a qualitative study learners' (the most critical figures of all assessment procedures) attitudes toward rejecting or retaining CLT principles were inquired. They, offering one caveat, which was the significance of teacher assessment to them, pronounced the method as a striking system.

© 2011 Elixir All rights reserved.

Introduction

Assessment refers to the systematic gathering of information for the purpose of making decisions or judgments about individuals. In all modern societies people undergo testing for both educational and non-educational purposes. Tests play a pivotal role in individuals' lives as some key decisions would be made based on the results. The traditional notion of testing has gone under significant changes after the commencement of the era of critical pedagogy and considering learners as a part of broad milieu of society, individuals whose voices and experiences and expectations should be validated in the educational systems. In the new way of testing which, in line with Critical Pedagogy, is called democratic assessment, learners are also participants and have their own "rights" that should be respected.

Shohamy (2001), one of the most striking figures in the field of Critical Language Testing, has proposed a comprehensive model counting CLT principles. In her model, she has attempted to enumerate most significant criteria for learners' voice to be heard. Her model has fifteen features that are presented shortly. Applying the critical perspective facets on assessment she has developed her inclusive CLT framework. The features that should be observed in a critical setting are as follows:

- 1- Critical language testing is not neutral, but is shaped by cultural, social, political, educational and ideological programs.
- 2- CLT encourages an active, critical feedback from test-takers.
- 3- CLT view test-takers as political subjects within a political

context. 4- CLT views tests as tools within a context of social and ideological struggle. 5- CLT asks questions about which and whose agendas tests serve. 6- CLT claims that testers need to understand the tests they create within a larger vision of society and its use of those tests. 7- CLT examines tests in terms of their measurement and assessment of knowledge vs. their definition and dictation of knowledge. 8- CLT questions the nature of knowledge that tests are based upon: whose knowledge? Independent 'truth' or negotiated and challengeable? 9- CLT examines the influence and involvement of the range of stakeholders in a testing context. 10- CLT perceives the embeddedness of tests within social and educational systems. 11- CLT admits to the limited knowledge of any tester and the need for multiple sources of knowledge. 12- CLT challenges the dominant psychometric traditions and considers 'interpretive' approaches to assessment that allow for different meanings and interpretations rather than a single absolute truth. 13- CLT considers the meaning of test scores within this interpretive framework, allowing for the possibility of discussion and negotiation across multiple interpretations. 14- CLT challenges the knowledge that tests are based upon and advocates a democratic representation of the multiple groups of society. 15- CLT challenges the primacy of the 'test' as assessment instrument and considers multiple procedures for interpreting the knowledge of individuals.

The present study will take advantages of three main principles of the above mentioned framework to run CLT guidelines in the language classrooms. These principles will be

Tele:

E-mail addresses: nrashidi@rose.shirazu.ac.ir

© 2011 Elixir All rights reserved

1- CLT encourages an active, critical response from test-takers. 2- CLT admits to the limited knowledge of any tester and the need for multiple sources of knowledge. And 3- CLT challenges the dominant psychometric traditions and considers 'interpretive' approaches to assessment that allow for different meanings and interpretations rather than a single absolute truth.

Objectives of the study

This study seeks to bring CLT principles into consideration and also make them practical in the actual learning milieu. It is hoped that with applying these rules, firstly both teachers and learners become aware of the "learners' right" in the learning and evaluation procedures. Secondly their passive roles transfer into the active and critical ones. Thirdly, the power distribute in a more balanced way amongst all individuals involved in the testing procedure.

Research questions

This study is conducted with the aim of answering the following questions:

- 1- Is there any difference in the evaluation of students based on both the traditional way of testing and the critical one?
- 2- In terms of consequential validity, can we continue the procedure of testing after the testing itself?
- 3- In what ways, if any, does CLT enhance learning?
- 4- Is it of worth to bring students' voices into the testing process?

Review of Related Literature

CLT is mainly rooted in the practice of critical pedagogy in the field of testing and evaluation. For this reason it would be of significance to take a short look at CP doctrine as well.

Critical Pedagogy

Critical pedagogy aims to raise learner's critical consciousness to be aware of their sociopolitical surroundings and to fight against the status quo, with the intent of transformation both in the classroom and in the larger society (Norton & Toohey, 2004). Critical pedagogy (CP) talks about the purpose and the process of education. Education in the critical sense lights up the connection among knowledge, authority and power (Giroux, 1994; cited in Steven, 2002). Brazilian educator Paulo Freire, the pioneering figure of the movement, initially talked about the issues of power and social injustice in the literacy programs and put forward critical pedagogy as an approach to face up with these issues (Baladi, 2007).

One of the distinctive features of critical pedagogy is dialogue. In the dialogical context, individuals in the classroom are considered as the members of one community in a way that all teach and all learn. In this manner, a mutual acceptance and trust between learners and the teacher would be created (Heaney, 1995). This approach of critical pedagogy is in contrast with the anti-dialogical method which holds a "hierarchical classroom structure of teacher over students" (Crawford, 1978, p. 91). Anti-dialogical approach sets the teacher as lecturer and the transmitter of knowledge and student as the depositor of knowledge (Freire, 1970).

Trying to define critical pedagogy in ELT, Canagarajah (2005) lends critical pedagogy to a practice-oriented stance where critical pedagogy "is not a set of ideas, but a way of doing learning and teaching. It is a practice motivated by a distinct attitude toward classrooms and society" (P. 932). Critical students and teachers are planned to bring their actual life experiences and needs to the classroom to demystify power implications in pedagogical activity and struggle to alter the

means and ends of learning in order to form more ethical, educational, and social environment. In this vein, Akbari (2008) contends that "the discourse of CP is the discourse of liberation and hope" (p. 277). Critical pedagogues in second language teaching are interested to explore the ways that social relationship and issues of power are settled in language, while often unnoticed in language studies and educational practices (Norton & Toohey, 2004). Norton and Toohey (2004) add that from this standpoint, language is not just as a means for communication rather it is "a practice that constructs, and is constructed by the ways language learners understand themselves, their social surroundings, their histories, and their possibilities for the future" (P.1). This CP is the base of Critical Language Testing that would be presented in next section.

Critical Language Testing

Critical theory is a tool of reason which can change the world used properly (Rasmussen, 1996). Critical Language Testing refers to questioning the use of tests as tools of power in educational settings (Shohamy, 2001). Shohamy believes that tests are tools of power and argues that tests are introduced by those at power centers in order to manipulate systems and impose their agendas on powerless test-takers. In this regard she supposes that test-takers are "true victims" of tests as they have to accept the authority of tests without questioning their contents and their relationships with their current level of knowledge. Bourdieu (1991) in the same vein criticizes the detrimental consequences of tests referring to the "symbolic power" of tests which is enhanced by encouraging test takers' dependency on tests and test results as the sole criterion of their evaluation from an early age.

Tollefson (1995) identifies the sources of power in all assessments by classifying them into three main categories: state, discourse and ideology. State power, he argues, is the bureaucracy of test; discourse power is the imposition of test by unequal individuals; and ideological power is about what is supposed to be right or wrong. Critical scholars refer to the unfavorable corollary of the power situations in which test-takers voice cannot be perceived in any way. The most significant element of CLT is test-taker as is identified as people in Taylor (1998). He talks about the right of the people; it is rule of the people, by people, and for people, and today "people" means everybody, without the implicit limitations of formerly excluded powerless people.

"Right of test-takers" is a new concept that has been introduced through the lenses of CLT. Traditionally test-takers were passive receivers of authorities' manipulation as Shohamy (2001) argues that test-takers were typically "black boxes" and were to be evaluated by some specific knowledge. She offered a 4-perspective framework with fifteen subcategories. The main themes of these principles are as follows: CLT should be considered in a broad setting of a society in which test-takers are political subjects within a political context. In CLT, there is a need for multiple sources of knowledge and interpretation of test scores. This model also encourages a democratic representation of multiple groups of society in constructing tests. Keesing-Style (2003) takes another aspect of CLT into reflection which is dialogic interaction. In this way "the role of teacher and student is shared and all voices are validated."

As it was made apparent the use of tests as instruments of power violates the fundamental principles of democratic assessment since the authority of tests is in the hands of powerful organizations that control knowledge (Shohamy,

2001). Fetterman, Kaftarian, and Wandersman (1996) argues that democratic assessment invite participants to examine the issues of concern in an open forum. In this way, they claim, "empowerment evaluation" would be brought into educational situations. In their study, Fetterman et al. discuss some forms of self-evaluation and reflection as tangible forms of the democracy in testing. Nevo (1996) also gives "dialoguing" as democratic assessment manifestation. Dialoguing is based on a two way communication between teacher and student and its basic assumption is that "nobody knows everything". In the same vein Freire (1985) rejects power models and supports democratic ones in which a meaningful dialogical relation exists between two partners: the evaluator and the evaluatee. Applying CLT principles, Shohamy (2001) believes, implies that testing cannot remain a territory belonging just to testers. Attitudes mentioned are all at the theory level. It should be also inspected if it is possible to bring these theories into practice or not, and if it is practical in all societies with different historical backgrounds. Guoxing Yu (2007) examined the use of students' voice in the evaluation of their written summaries and the result was to some extent contrary to the researcher's expectation. He reports the results as: it seems that the use of students' criteria for measurement was not so welcome and valued by the students who were accustomed to the common practice of using experts as defacto authoritative assessment criteria in China.

One of the important factors which play a critical role in all educational settings is the consequence of the testing method on the way that teaching materials are presented to the learners and the way students are supposed to learn. Washback effect or consequential validity which is relatively new concepts in the teaching and testing era have received a lot of attention in the last three decades. Bachman and Palmer (1996) consider washback to be a subset of a test's impact on society, educational systems and individuals. They believe that test impact operates at two levels: The micro level (i.e. the effect of the test on individual students and teachers); and the macro level or the impact the test may have on the society and the educational system. As G. Buck (1988) wrote, there is a natural tendency for both teachers and students to design their classroom activities to the demands of the test, especially when the test is very important to the future of the students, and pass rates are used as a measure of teacher success. This influence of the test on the classroom is, of course, very important; this washback effect can be either beneficial or harmful. In some situations, the way students get ready for the examination cause them to ignore the most important goal of language teaching which is communication not just success in the traditional paper and pencil tests.

Method

Participants

In the present study, 40 female students of English from a private language institute participated. They were in two classes with the same level of proficiency and the same teacher. Virtually, everything was parallel for them. Their classes were held two times a week, 2 hours each time. They were at intermediate level of proficiency.

Instruments

The instruments of the present study were some short, unfinished reading texts used as the writing topic which students were supposed to write an end to. Another instrument was a writing assessment scale evolved during the process of the study by the participants. This scale which will be discussed in more

details in the following sections has five elements including Idea, Grammar, Vocabulary, Cohesion/Coherence, and Punctuation. The traditional testing also makes use of Jacob's Rating scale.

Procedure

To give voice to test takers, and to divide the power among all entities in an assessment situation, and also to introduce a democratic way of testing, we exploited some CLT criteria. One factor was, as it was mentioned earlier, to encourage an active, critical response from test-takers. Birenbaum (1996) also believes that students should be active participants in the process of developing assessment procedures, including the criteria and standards by which performances are judged. For this reason, the learners who were the participants of this study had a very accentuated role in different phases of assessment. The full detail of the steps taken to reach this point will be discussed in the following section along with some other criteria. As CLT admits to the limited knowledge of any tester and the need for multiple sources of knowledge, we utilized multiple sources of assessment. In this regard learners themselves along with their teachers participated in assessing their writings. Students' voice would be heard as peer assessment as one of the actual forms of critical pedagogy will be added to the teacher's (the only source of knowledge in traditional testing) feedback. And finally, having the third principle in mind, which is: CLT challenges the dominant psychometric traditions and considers 'interpretive' approaches to assessment that allow for different meanings and interpretations rather than a single absolute truth, students' achievement was reported in the interpretative mode by both the teacher and other students. This is supported by other scholars such as Birenbaum (1996) who argues that reporting of assessment results should usually be in the form of a qualitative profile rather than a single score or other quantification.

Students' Developed Assessment Scheme

One of the critical facets of CLT based on what Shohamy (2001) believes is "the need to conduct and administer testing in collaboration and in cooperation with those tested" (p.376). Taylor (1998) mentions that "It is rule of the people, by the people, and for the people, and today the 'people' is taken to mean everybody, without the unspoken restrictions that formerly excluded peasants, women, or slaves" (p. 143). Following this principle as the main issue and as the first step to the democratic assessment, students developed their writing scheme based on which their papers would be assessed by the teacher and their peers. Posing the question "What would be important to you if you were to assess a written paper?" they thought profoundly, mainly for the first time, about their priorities. They were divided into three groups and discussed their values in assessment. They came to some brilliant guidelines (which surprised the researcher to some extent!) which were labeled by the help of the researcher. The first things that were of high significance to virtually all of the groups were vocabulary and grammar. They were asked to elaborate on what they mean by the meaning of these factors.

In a collaborative effort they discussed in more details what they believed was essential to them. In the "grammar" section, they cited the use of appropriate structure, the use of recently learned grammar, the use of the same structure in one piece of discourse as critical features. Regarding "vocabulary", the use of appropriate vocabularies (appropriate to their level of proficiency), correct spelling, correct capitalization, less use of all-meaning words such as thing, good, bad ,.....were mentioned.

A different issue noteworthy to them was that if the writer could talk about what she was going to say thoroughly or not, if the content is related to the topic, if the writer could develop and support the topic sentence or not. This feature was labeled as "Idea". Then they talked about the relationship between sentences. They believed that the sentences should be connected to each other in a meaningful way. They pointed to some connecting words such as also, but, however, although, on the other hand, and etc. As they were confused about the label to give to this feature, the label "cohesion/coherence" was suggested by the researcher (their teacher). The final point significant to learners was the importance of "punctuation". In their view, punctuation was a feature that could embellish and adorn the paper to the assessor. Some other points also were highlighted but they could be regarded as the sub-branches of these main features. So they came to this scheme based on the hierarchy of implication:

- 1-Idea: to pinpoint their true ideas and develop the topic.
- 2-Grammar: the use of appropriate structure, the use of recently learned grammar, the use of the same structure in one piece of discourse.
- 3-Vocabulary: the use of appropriate vocabularies (appropriate to their level of proficiency), correct spelling, correct capitalization, less use of all-meaning words such as thing, good, bad,....
- 4-Cohesion/Coherence: sentences should be connected to each other by cohesive devices and also in a logical way.
- 5-Punctuation: the appropriate use of punctuation (.,:;!)

These different labels were not weighted and all of them had the same value to them. This scheme was the baseline that the students used in other phases to reach a democratic evaluation in their classes.

Multiple Sources of Assessment: Peer and Teacher Assessment

In the traditional testing, the teacher is the only source of knowledge to the learners. As an authority, the teacher runs the engine of the assessment procedure. On the other hand, one of the main features of CLT as an alternative to the traditional way of assessment is that the teacher cannot and should not be the only source of knowledge. The limited range of the tester knowledge is admitted by CLT principles and the call for manifold sources of evaluation is occasionally underlined by CLT proponents. Shohamy (2001) among others argues that "in an interactive model the knowledge of different groups influence the knowledge of the dominant group and enrich it" (p. 384). She also points out that "CLT admits to the limited knowledge of any tester and the need for multiple sources of knowledge" (cited in Lynch, 2001. p.363). Freire (1985) pinpoint the issue in another word:

In understanding the process in this way, evaluation is not an act by which educator A evaluates educator B. It's an act by which educators A and B together evaluate an experience, its development, and the obstacles one confronts along with any mistakes or error. Thus, evaluation has a dialectical character . . . It's essential that members of the evaluating organization deeply believe that they have as much to learn from educators directly linked to popular bases as those who study at the bases. Without this attitude, the evaluators from an external organization will never admit to any gap between their view of reality and reality (p.23-25).

In the present study, in line with this principle, the power is distributed between the teacher and the student. The route of

evaluating the papers went through these stages. To prevent any sources of bias, the peers who were going to assess the papers were at the different class. The writings of the students were handed down to their peers studied at the same level of proficiency participating in the study as well. The participants are asked to read the papers and employing their own developed scheme, discussed in the previous section, assess their peers. The papers assessed in this way were turned back to the teacher for the second assessment. The same process is gone through by the teacher. Now the papers, with two different scores are handed down to their writers. They have their papers assessed by their teachers and their peers with some suggestive marks on them by the two. As the papers are marked with two sources of scores with the interpretive marks (discussed in the next section) attached to them, the learner could recognize a more crystal clear image of her level of proficiency. As the papers are assessed meticulously by the peers, there was a high correlation between the two scores ($R=0.76$).

Interpretive Mode of Assessment

One of the basis on which the traditional testing is standing and also it is criticized on that baseline is the reporting of one single score, mainly by the teacher, to the written papers. Critical testing criticizes this method and backs up the interpretive mode of assessment annexed to the scores. Shohamy goes for this way of assessment by mentioning that "CLT challenges the dominant psychometric traditions and considers 'interpretive' approaches to assessment that allow for different meanings and interpretations rather than a single absolute truth." (Mentioned in Lynch, 2001. p. 363). To this end, peers who were assessing the papers were to give their interpretive marks too. The continuum of very poor, poor, good, very good, and excellent described their interpretation. In general selecting just one of the labels of this continuum was not enough at all. They were to give their ideas about the paper in a few sentences as well. This way of scoring gives credit to their assessment and to the written paper at the same time. It is of significance to point out that, papers evaluated in the traditional mode lacks this property and may lead to the embarrassment of the students in interpreting their scores.

Till now, to practice the principles of CLT students' voices are evaluated through developing a scheme, assessing the peers' papers and also reporting the interpretive mode of scoring. But to be truly in line with critical way of assessment, our tests should not be stopped with reporting the scores in both modes of scoring and interpretive assessment. Referring to Messick's (1989, mentioned in Bachman, 1990) 'consequential validity', measurement and the validity of the test should go beyond the test situation itself. Having this principle in mind, we go a step further to include our testing outcome as a tool of teaching.

Test Consequences

To delve into the practice of the consequential validity in our study, we took advantages of two methods. The first one which is at the heart of the critical philosophy is "dialoguing" (Freire, 1985) and the second one is "the use of the learners' errors as the teaching materials". In the present study, the testee and the tester have the opportunity to participate in a dialogue to discuss the written paper's problems. So through the use of ongoing feedback the process of testing is not halted. It would be a dynamic process. So after receiving their papers, marked by two scores and the interpretive statements attached, the testee was encouraged to meet the teacher and her peer to talk about her problems. In this way through dialoguing, she could make an

enlarged image of her meaning and the interpretation of her test score. If the problems of the testees are entirely vanished from the subsequent performances or not is beyond the scope of this paper and is not examined. It needs to be completely scrutinized through another comprehensive study.

Bringing the learners' errors to the teaching arena is another effort to practice consequential validity in the study. In two ways the "noticing" procedure has been applied. The errors are divided into two categories. The first category involves the errors common to nearly all of the learners, such as agreement rule, the appropriate use of plural 's', and the rule of count/non-count nouns, and so on. They are used as teaching materials presented to the learners under the title of "your errors". They were taught directly and systematically every last 10 minutes of the class time. After teaching them, learners are to make as many sentences as possible with the correct structure. As it was mentioned earlier another study is needed to be conducted to observe if these problems disappear from the testees' future writings or not. The second category includes the errors related to the inappropriacy of using some words and some Penglish structures. These errors are noticed and taught when they arise in the future performances mostly in speaking and oral communication.

Traditional Testing

Another phase of the current study is the use of the traditional way of assessment. It should be noted that the term "traditional" does not refer to something obsolete and conventional that should be put aside and ignored in any cost. The term "traditional" here points out to the way of assessment which lacks the properties discussed so far. Having this in mind, we are not talking about the differences between scores when we are comparing traditional and critical testing.

In this regard, the participants were presented with another test of writing in this sense labeled as traditional assessment. The procedure is familiar to all of us. The topic was introduced, students wrote about two paragraphs about it, and the papers were handed down to the teacher to be evaluated and at last they were presented by a single score and the testing process came to an end. The papers were evaluated based on "Jacobs' Rating Scale" (presented in the Appendix) to reach a comprehensive score. The difference between the traditional and the critical testing, as it was mentioned before, is not because of the fact that we crave to make the scoring more precise or increase the testees' scores as the result of applying CLT principles.

As the tests are used to make decisions about individual test takers, they are the only group whose rights should be protected. As Shohamy (2001) argues, the concept of the rights of test-takers is relatively new. In this regard, she goes on, test takers have the right to question the use of the tests:

Test-takers should view it as their right to question tests and the values inherent in them. Test-takers can also question the test results and methods wherever there is a feeling that the rights of the test-taker have been violated. They may have been tested on unfamiliar material, using unfamiliar methods; the test results may be used for purposes for which they were not intended.

So we come to the point that test takers as critical figures of testing process have the right to reject the present way of testing or decide to continue with it. For this reason their attitudes ought to be of high value to testers and testing institutions.

What about Testees? Are They into CLT Practice?

As it was thrashed out in the previous section, it should be the test takers who must decide finally to reject or retain a

system of assessment. To this end we were to conduct a comprehensive interview with all learners. Unfortunately we faced some practicality obstacles prevented us to do so. For this reason, the interview questions were written down and the testees were to answer them. In the questions, the writing based on the traditional way of assessment was referred to as Test A, and the test in line with critical assessment was called Test B. In a question, students were asked if they could choose one of the tests (A or B) which one they would choose. Nearly all students were into Test B for the reasons like these:

- It has two learning advantages: for both the assessee and the assessor; peers also could evaluate themselves.
- Writing by Test B opens my mind into using new vocabularies used by the peers in my future writings.
- In this way I can assess myself.

They mostly believed that this was a way to improve their own learning in a positive way. In another question they were asked about the two scores they were presented with. Looking at it through rose color glasses they argued that it was of high significance to them but only when the peer's score is supported by the teacher's score:

- That we were evaluated by two different scores is very informative, but teacher's idea is more important to me.
- I think besides the peer's score, teacher's score makes the score more reliable.

In discussing the interpretive way of assessment, they had some contradictory ideas:

- "By qualitative assessment, which is replacing the traditional quantitative one, we have a range for our scores not just a single score. A single score cannot represent students' abilities".
- Descriptive scores are very important to describe the students' level but they should be added to the quantitative score not to replace it.
- By my score supported by a description, I know where I am exactly standing.
- This report is advantageous to a quantitative score as it decrease the acute competition in educational settings.

In a complementary question, they were asked about the practicality of using learners' attitudes in the evaluation process. Nearly all of them had the same viewpoints: Surely but if supported by teacher's attitude. Asked about the advantages of test B, they replied:

- In this method, everybody is active and creative.
- It brings our own problems into the frontline.
- Developing the scale taught me to think more deeply about my learning.
- I was always afraid of being tested. It made me more relaxed and stress free.
- I have a feeling that we also can decide about our learning and evaluation.
- I try to take advantages of others' description about my work.
- With two assessment and descriptive report, I have a clearer image about my level of proficiency.

With this question, learners could eloquently describe the positive points of their voices to be heard through testing. Pointing to the teaching of the students' problematic areas following the test, they believed that we could make use of Test B mostly as a way of teaching besides assessment.

On the whole we can argue that testees as the central figures in all educational curriculums can decide about the utility of ways they are going to be assessed with.

Results and Discussion

The study was an attempt to exploit three features of CLT principles to validate students' voices in line with democratic society characteristics. In this section we attempt to answer our research questions posed earlier by using both qualitative and quantitative data analysis.

Quantitative Analysis

To discuss the first question: is there any difference in the evaluation of students based on both the traditional way of testing and the critical one? We need to argue that the distinction between the traditional and the critical testing, as it was mentioned before, is not because of the fact that we want to make the scoring more precise or increase the testees' score as the result of applying CLT principles. As we expected beforehand, running the ANOVA test (Table 1) showed that there was no significant difference between CLT and traditional assessment scores. The only difference was in the way that power was distributed among all individuals involved in the testing process and the knowledge was out of the dominant authorities' control to other members as well, and at the same time learners' voices could be heard. By this kind of interpretation we can support our practice as a valid and conclusive evaluation although on the face of it there would be no difference in the short run achievement of the learners. As Freedman (1993) also argues for working collaboratively with schools and communities, involving teachers, administrators, students and parents to come along CLT principles, we made use of learners' assessment to be on CLT road.

As another quantitative part of the study, we go to make an eloquent and cogent answer to the fourth question: Is it of value to bring students' voices into the testing process? The answer is definitely yes. Fetterman *et al.* (1996) made use of a form of self-evaluation and reflection in which the evaluators act as facilitators and collaborators rather than as experts and counselors. They claim that such an approach is fundamentally democratic as it invites participation through which issues of concern to the entire community are examined in an open forum (cited in Shohamy, 2001). As another device we made use of peer assessment as a democratic assessment. Before going into the details, I need to mention that we argue this question from both quantitative and qualitative analysis. From the quantitative point of analysis, we believe that, besides its learning consequences (discussed in the following parts), we distributed power among all people engaged by participating students in evaluating themselves not just being evaluated. In this regard we can refer to the high correlation between students' and teachers' scores assigned to one specific paper. The correlation may make students' scores as a valid indicator of the scores they reported. Table 2 shows this correlation.

It worths mentioning here that it is not because of this correlation that we would make use of students' voice. They would participate in the process to make testing an emancipatory practice. The high correlation is just to appease our worriness towards the validity of the scores that students report.

Qualitative Analysis

Questions 2 and 3 are going to be discussed qualitatively. These questions can be answered based on two sections of the study; the first one with the reference to the consequential validity we discussed earlier and the second one in reference to students' attitudes toward CLT. The process that we took advantages of to utilize learners' mistakes into teaching was discussed in the procedure section eloquently. Lynch (2001)

pointing to the importance of consequential validity in critical perspective argues that "the question to be examined is: what specifically is done as a result of the assessment (e.g., is a change made in the curriculum; does a teacher-in-development alter some aspect of his or her teaching style)?" Just to avoid repeating what went there, I just touch upon its values. The value of bringing learners problematic areas into teaching milieu has been supported through ample teaching researches. In the present study, also we didn't stop after the test itself and went beyond to a diagnostics level. Under two different titles the mistakes were divided and discussed separately in the classroom. Although we can be cocksure that it is of very high significance to continue the testing process to the teaching one, we cannot claim that the mistakes, taught systematically, would be disappeared from learners' consequent performance or not. This issue can be discussed more precisely referring to students' attitudes. If you take a short look at part 3.3.6, you may come to the conclusion that the value of CLT for learners is mostly summarized with regard to its learning consequences. They generally discussed the virtues of CLT with pointing to what they could learn in this process to make them autonomous writers. What is apparent in the analysis of learners' attitudes verifies the research done by Yu (2007) in an Eastern culture: "the use of a popular scoring template became, unfortunately, a kind of imposed 'democracy' on the students who were accustomed to the common practice of using experts as de facto authoritative assessment criteria and were more than willing to maintain the current practices" (p.560). Participants in our study in the same vein, although cherish the democratic assessment, look at the teacher as an authoritative figure in the class.

Concluding Remarks

As it has been argued by critical testing scholars, testing power has been long kept in the authorities' hands in a hierarchical way, from political organizations at the top and classroom's teacher at the bottom. Learners, however, although the most important individuals in the testing process were ignored completely. They were, and still are, manipulated by the figures at the power centers.

In this study we attempted to distribute the power, traditionally hold by teachers as the only source of information, in a more unprejudiced way between teachers and learners. As the first step towards this power allotment, learners developed a writing assessment scheme collaboratively with their teacher not as a super ordinate figure any longer.

The scheme was used in the consequent steps (peer assessment) as a valid scheme. Another way, to this end, was the use of peer assessment along with teacher assessment supported by the interpretive mode of assessment. In this way we admitted to the limited knowledge of the teacher in assessing learners and asked for the learners' voice in the evaluation process. As the last point, students' problematic areas, reported by both the teacher and peers, were not left behind. They were congregated and under the title of "your mistakes" through noticing process taught in a systematic way. As the critical figures of any testing situation, learners pinpointed their positive and negative points towards critical language assessment to be used in the future plans.

Although presented so rigorously, CLT principles as an emancipatory evaluation method are not practiced in the educational arena. Instead, the practice of the traditional testing with all its limitations has dominated our learners and muddled the road into a broader, democratic assessment.

References

1. Bachman, L.F. 1990: Fundamental considerations in language testing. Oxford: Oxford University Press.
2. Bachman, L. & Palmer, A. (1996). Language testing in practice. Oxford University Press.
3. Birenbaum, M. (1996). Assessment 2000: Towards a pluralistic approach to assessment. In Birenbaum, M. and Dochy, F.J.R.C., editors, Alternatives in assessment of achievements, learning processes, and prior knowledge. Dordrecht, Netherlands: Kluwer, 3–29.
4. Bourdieu, P (1991). Language and symbolic power: Cambridge, MA: Harvard University Press.
5. Buck, G. (1988). Testing listening comprehension in Japanese university entrance examinations. *JALT Journal*, 10, 15-42.
6. Canagarajah, S. (2005). Critical pedagogy in L2 learning and teaching. In E. Hinkel (Ed.) *Handbook of research in second language teaching and learning* (pp. 931-949). Mahwah, NJ: Lawrence Erlbaum Associates
7. Crawford, L. M. (1978). Paulo Freire's philosophy: derivation of curricular principles and their application to second language curriculum design. Unpublished doctoral dissertation, University of Minnesota. Retrieved from ProQuest Digital Dissertations. (AAT 7911991)
8. Fetterman, D., Kaftarian, S. & Wandersman, A. (1996). Empowerment evaluation. Thousand Oaks, CA: Sage.
9. Freedman W.S. (1993). Linking large-scale testing and classroom portfolio assessments of student writing. *Educational Assessment* 1, 27–52.
10. Freire, P. (1970). *Pedagogy of the oppressed* (M. B. Ramos, Trans.). New York: Continuum.
11. Freire, P. (1985). *The Politics of Education*. Hadley, MA: Bergin and Garvey.
12. Heaney, T (1995). Issues in freirean pedagogy. Retrieved May 30, 2010, from <http://www3.nl.edu/academics/cas/ace/resources/Documents/FreireIssues.cfm>
13. Keesing-Style, L. (2003). The relationship between critical pedagogy and assessment in teacher education. *Radical Pedagogy*, 5(1). Retrieved Oct 22, 2010, from http://radicalpedagogy.icaap.org/content/issue5_1/03_keesing-styles.html
14. Lynch, B. K. (2001). Rethinking assessment from a critical perspective. *Language Testing*. 18, 351-372.
15. Nevo, D.)1996(. *School based evaluation*. Oxford: Pergamon University Press
16. Norton, B., & Toohey, K. (2004). Critical pedagogies and language learning: an introduction. In B. Norton & K. Toohey (Eds.), *Critical pedagogies and language learning* (pp. 1-17). Cambridge: Cambridge University Press.
17. Rasmussen, D.M. (1996). Critical theory and philosophy. In Rasmussen, D.M., editor, *The handbook of critical theory*, Oxford: Blackwell, 11–38.
18. Shohamy, E. (2001). Democratic assessment as an alternative. *Language Testing*. 18, 373-391
19. Shohamy, E.(2001). *The power of tests: a critical perspective on the uses of language test* London: Longman
20. Steven, CH. (2002). Critical pedagogy on the web: Key terms and concepts related to critical pedagogy and educational theory and practice. Retrieved Oct 20, 2010, from <http://mingo.info-science.uiowa.edu/~stevens/critped/terms.htm>
21. Tollefson, J. (1995). Introduction: language policy, power, and inequality, in Tollefson, J., editor, *Power and inequality in language education*. London: Cambridge University Press
22. Taylor, C.)1998(. *The dynamics of democratic exclusion*. *Journal of Democracy*. 9, 143– 56.

Table 1. ANOVA results for traditional and critical testing scores

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	143.333	2	71.667	.176	.839
Within Groups	23200.000	57	407.018		
Total	23343.333	59			

Table 2. The correlation between teacher and peer assessment

		Peer	Teacher
Peer	Pearson Correlation	1	.762**
	Sig. (2-tailed)		.000
	N	38	37
Teacher	Pearson Correlation	.762**	1
	Sig. (2-tailed)	.000	
	N	37	37