Alternative assessment or traditional testing: How do Iranian EFL teachers respond?

Abbas Ali Rezaee
Associate Professor, University of Tehran, Iran
Seyyed Mohammad Alavi
Associate Professor, University of Tehran, Iran
Enayat Shabani[1]
Ph.D. Candidate, University of Tehran, Iran

Abstract

Introducing alternative modes of assessment is but one response to the recent call for democratic and ethical language assessment. Yet, despite the recent emphasis in the discourse community and the rise in publication on alternative assessment, these new forms of assessment still need to be explored further. This study is a two-fold attempt: first, to investigate teachers' attitudes and beliefs about different aspects of traditional testing and alternative assessment, and second to delve into their ethical orientation and to examine views on language testing apropos of their general ethical viewpoints. A questionnaire was developed and used to collect Iranian EFL teachers' views on language testing and ethics in general ($N = 153$). The results indicated that despite its agreed-upon disadvantages, an obstinate stigmatization and refusal of traditional testing may still seem a practice at odds with the common sense. In fact, until a better proposal can be offered, alternative assessment and traditional testing can best be regarded as supplements rather than substitutes.

---

[1] Corresponding author: en_shabani@yahoo.com

Rezaee, Alavi, and Shabani

1. Background

The recent call for democratic and ethical language assessment (e.g. Shohamy, 2001a) has been taken in different ways. Developing a professional body of knowledge (e.g. Davies, 1997; Stansfield, 1993), introducing alternative modes of assessment (Brown & Hudson, 1998; Hamayan, 1995), examining, questioning, and redefining validity and reliability (e.g. Messick, 1980, 1996; Lynch, 2001a; Moss, 1992), involving stakeholders in language assessment (e.g. Mathew, 2004; Rea-Dickins, 1997), considering language assessment as social practice (e.g., McNamara, 2001a; McNamara & Roever, 2006), taking the consequences and power of tests on board (e.g. Alderson & Wall, 1993; Hamp-Lyons, 1997; Shohamy, 1997, 2001b; Wall & Alderson, 1993), and developing new methods for detecting and removing bias in tests (Elder, 1997; Holland & Thayer, 1988; Holland & Wainer, 1993; Penfield & Camilli, 2007) are but some responses to this call, but in Lynch's (1997) eloquent words, "the search for ethical assessment continues" (p. 324).

Tests have always been given to students as a means of estimation of competence (Hancock, 1994). For the meantime, students seem to be fully conversant with the institutional conventions of traditional testing. In fact, it is not implausible to assume that almost any student who finishes their secondary education has, at least once in their life, taken a test after which they were given a single score. At times, important decisions are made based on the scores: Students pass, fail, or retake a course; teachers get approved or reproved; teaching materials will change; and even instructional methods and techniques may get revised.

Alternative forms of assessment are believed to represent a new promising generation of language tests to address ethical concerns in the field. The increasing dissatisfaction with traditional testing techniques encouraged educational assessment professionals to think of some new (or alternative) methods to overcome different stakeholders' discontent with traditional tests. Alternative assessment includes the more student-centered forms of assessment, such as portfolios, interviews, journals, project works, and self or peer assessment which are essentially different from the

conventional forms of testing. In fact, in contrast to traditional testing techniques, this new method concentrates on the process of learning and uses assessment as a means at the service of promoting student learning. Consequently, it is maintained that alternative assessment has positive washback on education (Hamp-Lyons, 1997).

Usually carried out constantly over a period of time, these forms of assessment primarily focus on assessing learners' overall and holistic ability in utilizing language in real-life situations. To Hamayan (1995), alternative assessment "refers to procedures and techniques which can be used within the context of instruction and can be easily incorporated into the daily activities of the school or classroom" (p. 213). Alderson and Banerjee (2001) believe that in comparison to traditional testing, alternative assessment is less formal, and is usually formative rather than summative in function.

Language testers have generally come to recognize the limitations of traditional methods of estimating what students know and what they have learnt. In fact, they have begun to consider more innovative ways for this purpose, which are now generally subsumed under the alternative assessment umbrella, though they may take different forms with noticeable differences in details. Alternative forms of assessment are believed to represent a new promising generation of language tests to address the growing concerns in the field. Alternative assessment includes the more student-centered forms of assessment, which are essentially different from the conventional forms of testing. In fact, in contrast to traditional testing techniques, the new methods concentrate on the process of learning and use assessment as a means at the service of promoting student learning. In fact, that is why alternative assessment is believed to have positive washback on education (Hamp-Lyons, 1997).

Hamayan (1995) holds that the information provided through alternative assessment procedures can be of value for different groups of stakeholders. Teachers can assess students in an appropriate natural setting and can obtain data throughout the course on the strengths and weaknesses of their curriculum, and find the opportunity to revise and modify it at any time of the course based on their students' accomplishments and their ongoing

Rezaee, Alavi, and Shabani

requirements. Students, too, can see their own progress in a less daunting setting, yet not in an esoteric language made by specialists to be used only by specialists, and they can assume more responsibility for their own learning (Hassel & Lourey, 2005). Moreover, parents can get a clear picture of their children's performance, too. Finally, administrators, who are described as the "least convinced of the advantages of alternative assessment" (Hamayan 1995, p. 215), have also found alternative assessment useful in different aspects (see Clapham, 2000, Tsagari, 2004).

2. Purpose of the Study

Despite the recent emphasis in the discourse community and the rise in publication on alternative assessment, these new forms of assessment still seem to have a long way to go before they can reach to their full maturity. Yet, the purpose of the present study is not to prescribe any ethical standard to language testers or to get to any unanimous conclusion on the ethical policy to be pursued in the field. Prior to solving any problem, the problem should be understood and described. By taking a postmodern perspective, which acknowledges people's diversity of perception and defines reality in relation to who the interested parties are and what their interests hinge on, the present study is an attempt to bring some of the concerns in the field to the front. In particular, it seeks to examine the two research cultures or paradigms (Lynch, 1997) in the field (i.e., language testing and language assessment), especially in the eyes of the major stakeholders. The study, in fact, is a two-fold attempt: first to investigate teachers' attitudes and beliefs about different aspects of traditional testing and alternative assessment, and second to delve into their ethical orientation and to examine views on language testing apropos of their general ethical viewpoints.

3. Method
3.1 Participants

A questionnaire was used to collect the teachers' views on language testing and ethics in general (see Appendix). It was sent

electronically to 326 English teachers from whom 153 filled the questionnaire (N = 153). The respondents were adult EFL teachers, 56% male and 42% female. About 2% did not reveal their gender. Almost 21% of them expressed having one to five years of teaching experience, 46% had between six to ten, and 33% had more than 11 years of teaching experience. Of the English teachers who participated in the study, 95% had a degree in TEFL. Their ages ranged from 20 to 59, though the majority of them (82%) were between 20 to 39 years of age. All the participants held, at least, a master's in their major.

3.2 Instrumentation, Data Collection, and Data Analysis

Being inductive in approach, qualitative description is a qualitative method which can be used for identifying a problem, generating hypotheses, formation of theories, and developing concepts. Aiming to "describing the informant's perception and experience of the world and its phenomena" (Neergaard, Olesen, Andersen, & Sondergaard, 2009, p. 2), one of the primary uses of qualitative description is in developing questionnaires.

In the present study, semistructured individual interviews with open-ended questions were used (n = 9). The issues suggested by Sandelowski (2000) in designing qualitative description (i.e., philosophy, sample, data collection, analysis, and outcomes) were taken into account. Also, the common analytic strategies for content analysis in qualitative description (see Miles & Huberman, 1994) were followed. Efforts were made to consider the strategies to enhance the rigor of description in terms of authenticity, credibility, criticality, and integrity (Milne & Oberle, 2005). Data from the interviews were coded, insights and reflections on the data were recorded, similar patterns and themes, similarities, and differences among the data were identified and extracted for further consideration and analysis; finally, prudent generalizations were made and then compared against the existing knowledge.

A pool of 71 items was developed and the first draft of the questionnaire was prepared. Two experts in language testing and a nonspecialist in the field (following Dörnnyei, 2003) were asked to go through the questionnaire and answer it while focusing on

wording, meaning, clarity, and relevance of the items in the presence of one of the researchers, so that the respondents' reactions and verbal comments could be taken note of. At the end of the draft, there was a section for them to write their possible comments or suggestions. Also, they were asked to write any other question(s) that they thought should or could have been incorporated in the questionnaire.

The questionnaire was then sent, electronically, to 38 English teachers from among whom 22 responded and returned the form online. After a preliminary analysis of the results and discussion with experts, the items were reduced to 33 categorized under four parts. The questionnaire had a general instruction, with specific instructions for each group of questions. As much of the final version did not change after the second piloting, and as the respondents were not informed of the piloting process, the data from the second piloting were used in the main study for most respondents.

The questionnaire was designed and made available online, and an invitation letter was sent to the participants electronically, including a hyperlink to the questionnaire Web page. In an effort to apply an electronic snowball sampling procedure, and also to remove the problem of respondent self-selection (see Bethlehem, 2008 for a complete discussion on this problem), some of the respondents kindly accepted to forward the questionnaire to other EFL teachers they knew. All in all, 153 English teachers filled and returned the questionnaire, showing a response rate of about 42%. Relying on the numbers reported by proxies, altogether 343 teachers received the questionnaire. It should be noted that 22 teachers had already completed the questionnaire in the piloting stage.

On submitting the questionnaire, the respondents received a thank you message including the e-mail address of one of the researchers. Interested respondents could receive a summary of the findings by sending an e-mail expressing their interest.

4. Results and Discussion

The data from the questionnaire were coded and entered into a computer file using SPSS. Other computer programs were also used

to analyze the data including ViSta-PARAN (Young, 1996) and MonteCarlo PA (Watkins, 2000). Usual checks were run to remove contradictions, impossibilities, and implausibility from the data. Also, principal component analysis (PCA) was run on Part I which contained 27 items of the questionnaire.

4.1 Principal Component Analysis of Part I (Agree-Disagree Opinion Questions)

PCA (originally proposed by Hotelling, 1933) was used to analyze the data. Before the PCA was conducted, its assumptions were checked. As the Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO) measures or Bartlett test were not produced in the output of the analysis, yet at the same time, the results of the PCA proper were produced, it was implied that the correlation matrix was positive semidefinite and not positive definite, as it is usually expected to be by default in programs such as SPSS. The implication was that there was, at least, one 0 eigenvalue, which was verified by finding five 0 eigenvalues in the output. Nonetheless, as the PCA could be applied, and the output was produced, it could be concluded that there would be no negative eigenvalues.

In fact, as the determinant of the matrix is 0 in such cases, and the matrix cannot be inverted using a standard inverse and is, thus, singular, the KMO values cannot be computed as the analysis requires an inversion of the matrix. Also, because the determinant of the matrix was 0, the Bartlett test statistic could not be produced either, as it is a function of the log of the determinant and the log of 0 is undefined. The correlation matrix for the variables contained several correlations of 0.30 or greater.

There are different rules used and discussed in the literature for deciding the number of principal components to retain. The simplest technique is to examine the PCA output for the total variance and select a cumulative percentage of total variance which one desires to be accounted for by principal components. The smallest number of principal components that explain the desired percentage will, then, be the number of Principal Components (PCs) to retain. The usual practice is to set the cut-off cumulative percentage between

Rezaee, Alavi, and Shabani

70% to 90%. Using this technique, seven principal components could be retained in the present study.

Another way to decide the number of PCs to retain is Kaiser's criterion or rule (also known as eigenvalue rule or K1 method). Here, PCs with an eigenvalue of, at least, 1.0 are retained and considered for further examination. In the present study, nine components were found to meet this criterion. Altogether, 80.07% of the variance was found to be explained by these nine components. The unrotated loadings of each of the items on the retained components were extracted. By default, K1 method is used to retain components with eigenvalues more than 1. Although nine components were extracted, very few items loaded on component 9, suggesting an eight-factor solution. The rotated nine-factor solution was also examined, and the nine-factor solution was found less optimal than the eight-factor solution.

In addition, Monte Carlo PCA for Parallel Analysis (Watkins, 2000) was used, and the first eight values from the principal component analysis were found to be larger than the values obtained from parallel analysis. In order to visualize the parallel analysis output, a plug-in of The Visual Statistics System (AKA ViSta; Young, 1996) was used. This plug-in, which is developed with LipStat statistical programming language, is called ViSta-PARAN (Young, Ledesma, Molina, & Valero, 1997). Using normal data simulation, 999 samples were simulated (cut-off percentile = 95th). The visual output consisted of a scree parallel plot (on the left in Figure 1) and a scree simulation plot (on the right in Figure 1). The observed eigenvalues and the 95th percentile random data eigenvalues intersect at the eighth principal component, suggesting the retention of eight PCs:

Alternative assessment or traditional testing
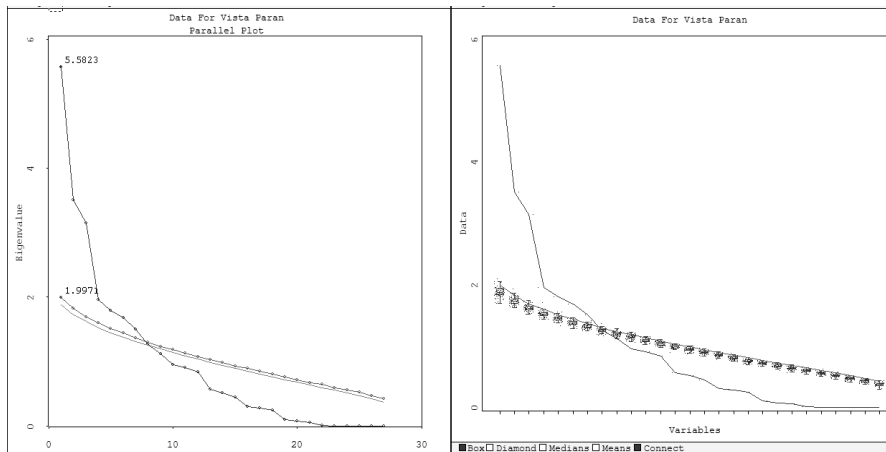
Figure 1: The visual output of ViSta-PARAN (cut-off percentile = 95th; 999 simulated samples)

Eight-factor solution: An eight-factor solution was implemented, and it was found that %76 of the variance could be explained by this solution. The results from Oblimin with Kaiser Normalization indicated that the main loadings on Component 1 were items 24, 7, 14, and, 10 generally referring to practicality, administrability, and exactness of alternative assessment. The main items on Component 2 were items 9 and 8, both referring to the beneficial quality or utility of alternative assessment. The main loadings on Component 3 were found to be items 18 and 5 which address the context and situation to use alternative assessment and traditional testing. Item 26 was found to have the highest loading on Component 4, addressing the necessity of including and using both alternative assessment and traditional testing in classroom practice.

The highest loadings on Component 5 were items 15, 2, 3, 4, and 6 which generally address different qualities of traditional testing including adequacy in measurement of learning, administrability, its benefits, and its ethicality. Items 22 and 23 had the highest loadings on Component 6 and concerned the difficulty of administration of alternative assessment at university level and emphasized its supplementary function to traditional testing. Items 1 and 12 had the highest loadings on Component 7 addressing teachers' satisfaction with the current practices and approaches and

Rezaee, Alavi, and Shabani

the ethicality of the present approaches to language testing and evaluation. The highest loadings on Component 8 were found to be items 13 and 19 which address objectivity and exactness of traditional testing in demonstrating an objective picture of students' progress in a given course. Moreover, from the Communalities obtained from the eight-factor rotation solution, it was found that all the items fit well with other items in the components. Also, the scale was found to have good internal consistency, with a Cronbach alpha coefficient of .78.

4.2 Analysis of Items: Findings and Discussion

It should be noted here that as the items of the questionnaire were put in a random order not to lead the respondents into certain answers, the analysis that comes below does not follow the same order as items appeared in the questionnaire.

About 95% of the respondents agreed that "traditional testing is easy for correction" (item 2), and none of the respondents expressed disagreement with the statement. A mean of 1.86 for all the respondents indicates a strong agreement with the statement. Conversely, none of the teachers strongly agreed that "alternative assessment is easy for correction" (item 7). In fact, 73% of them either strongly disagreed or disagreed with the statement. A mean of 5.63 indicates a general disagreement on the part of the teachers with alternative assessment being easy for correction (Tables 1 and 2 below).

Table 1: Frequency of responses for item 2

		Frequency	Valid Percent	Cumulative Percent
	1	49	32.0	32.0
	2	83	54.2	86.3
Valid	3	14	9.2	95.4
	4	7	4.6	100.0
	Total	153	100.0	

Table 2: Frequency of responses for item 7

		Frequency	Valid Percent	Cumulative Percent
Valid	2	7	4.6	4.6
	3	7	4.6	9.2
	5	28	18.3	27.5
	6	90	58.8	86.3
	7	21	13.7	100.0
	Total	153	100.0	

None of the respondents strongly agreed with the statement that "Through traditional testing, teachers can provide students with ample feedback on their progress and performance throughout the course" (item 4). It was found that 46% of them either strongly disagreed or disagreed with the statement (Table 3). A mean of 5.07 and a mode of 7 suggest a general disagreement with the statement. However, the teachers found alternative assessment practical in this regard, and, in fact, 36% of them strongly agreed with the statement that "Through alternative assessment, teachers can provide students with ample feedback on their progress and performance throughout the course" (item 9). Also, 79% of them agreed with the statement. Nobody strongly disagreed and only 14% expressed disagreement.

Rezaee, Alavi, and Shabani

A mean of 2.19 for all the responses to this item suggests a general agreement among the teachers with this statement (Table 4).

Table 3: Frequency of responses for item 4

		Frequency	Valid Percent	Cumulative Percent
Valid	2	18	11.8	11.8
	3	14	9.2	20.9
	4	21	13.7	34.6
	5	29	19.0	53.6
	6	28	18.3	71.9
	7	43	28.1	100.0
	Total	153	100.0	

Table 4: Frequency of responses for item 9

		Frequency	Valid Percent	Cumulative Percent
Valid	1	56	36.6	36.6
	2	65	42.5	79.1
	4	18	11.8	90.8
	5	7	4.6	95.4
	6	7	4.6	100.0
	Total	153	100.0	

About 58% of the teachers disagreed with the statement that "Alternative assessment is administrable in almost all situations and classes" (item 10), whereas only 32% of them agreed with it (Table 5). A mean of 4.39 and a mode of 5 for all the responses on this item suggest a general disagreement with the statement. However, as can be seen in Table 6 below, 79% of them agreed that "Alternative assessment is beneficial to learners" (item 8). A mean of 2.56 suggests a general agreement among the participants with the statement. Yet, nobody strongly agreed with it, and only 9%

agreed with the statement "Traditional testing is beneficial to learners" (item3) (see Table 7 below). A mean of 4.13 for all the participants points to a mild disagreement with the statement. However, as seen in Table 8, more than 65% of the teachers believed that "Traditional testing is administrable in almost all situations and classes" (item 5).

Table 5: Frequency of responses for item 10

		Frequency	Valid Percent	Cumulative Percent
Valid	1	14	9.2	9.2
	2	28	18.3	27.5
	3	7	4.6	32.0
	4	15	9.8	41.8
	5	35	22.9	64.7
	6	32	20.9	85.6
	7	22	14.4	100.0
	Total	153	100.0	

Table 6 : Frequency of responses for item 8

		Frequency	Valid Percent	Cumulative Percent
Valid	1	42	27.5	27.5
	2	64	41.8	69.3
	3	15	9.8	79.1
	4	4	2.6	81.7
	5	7	4.6	86.3
	6	21	13.7	100.0
	Total	153	100.0	

Rezaee, Alavi, and Shabani

Table 7: Frequency of responses for item 3

		Frequency	Valid Percent	Cumulative Percent
Valid	2	14	9.2	9.2
	3	50	32.7	41.8
	4	33	21.6	63.4
	5	14	9.2	72.5
	6	42	27.5	100.0
	Total	153	100.0	

Table 8: Frequency of responses for item 5

		Frequency	Valid Percent	Cumulative Percent
Valid	1	28	18.3	18.3
	2	43	28.1	46.4
	3	29	19.0	65.4
	4	25	16.3	81.7
	5	21	13.7	95.4
	6	7	4.6	100.0
	Total	153	100.0	

The results on "Exact results can be expected from traditional testing" (item 13) will be more significant juxtaposed with those on "Exact results can be expected from alternative assessment" (item 14). About 26% of the respondents selected option 2 on item 13, denoting moderate agreement, whereas 27% selected moderate disagreement. On item 14, about 27% of the participants selected option 2, and 27% selected option 6. About 49% of the participants agreed and 42% disagreed with the statement in item 13 (nobody strongly disagreed). On item 14, about 42% expressed their agreement and about 44% disagreed. The mean for all the responses on these two items were found to be 3.85 and 4.12, respectively,

suggesting a somewhat neither-agree-nor-disagree position on both items (Tables 9 and 10).

Table 9: Frequency of responses for item 13

		Frequency	Valid Percent	Cumulative Percent
Valid	1	7	4.6	4.6
	2	40	26.1	30.7
	3	28	18.3	49.0
	4	14	9.2	58.2
	5	22	14.4	72.5
	6	42	27.5	100.0
	Total	153	100.0	

Table 10: Frequency of responses for item 14

		Frequency	Valid Percent	Cumulative Percent
Valid	2	42	27.5	27.5
	3	22	14.4	41.8
	4	21	13.7	55.6
	5	19	12.4	68.0
	6	42	27.5	95.4
	7	7	4.6	100.0
	Total	153	100.0	

Whereas about 28% disagreed with the statement that "Alternative assessment adequately measures the learning outcomes" (item 16), 55% of the respondents agreed with it (Table 11). Taking all the responses on this item into account, a mean of 3.4, however, suggests a feeble inclination among all the participants to agree with the statement. About 65% of the teachers expressed their disagreement with the statement that "Traditional testing adequately measures the learning outcomes" (item 15). In fact, nobody strongly agreed with the statement (Table 12). The mean for all the responses on this item was found to be 4.91,

Rezaee, Alavi, and Shabani

suggesting a general disagreement with the statement. As can be seen in Table 13, it was also found that 95% of the participants expressed their agreement with the statement "Alternative assessment forms an essential part of education for its flexibility and adjusting to the student learning styles and individual development" (item 17). A mean of 2.07 for all the responses on this item indicates a general agreement among all the respondents.

Table 11: Frequency of responses for item 16

		Frequency	Valid Percent	Cumulative Percent
Valid	1	14	9.2	9.2
	2	42	27.5	36.6
	3	28	18.3	54.9
	4	26	17.0	71.9
	5	21	13.7	85.6
	6	22	14.4	100.0
	Total	153	100.0	

Table 12: Frequency of responses for item 15

		Frequency	Valid Percent	Cumulative Percent
Valid	2	14	9.2	9.2
	3	25	16.3	25.5
	4	14	9.2	34.6
	5	29	19.0	53.6
	6	50	32.7	86.3
	7	21	13.7	100.0
	Total	153	100.0	

Alternative assessment or traditional testing

Table 13: Frequency of responses for item 17

		Frequency	Valid Percent	Cumulative Percent
Valid	1	35	22.9	22.9
	2	86	56.2	79.1
	3	25	16.3	95.4
	5	7	4.6	100.0
	Total	153	100.0	

Of all the participants, about 91% disagreed with the statement that "Alternative assessment should be used in primary and secondary education (and not higher education, i.e. at university level)" (item 18). A mean of 5.7 and a mode of 6 suggest a general disagreement with the statement among all the respondents. It can be argued that the teachers considered alternative assessment an essential part of higher education as well as primary and secondary education (Table 14).

Table 14: Frequency of responses for item 18

		Frequency	Valid Percent	Cumulative Percent
Valid	2	7	4.6	4.6
	3	7	4.6	9.2
	5	26	17.0	26.1
	6	85	55.6	81.7
	7	28	18.3	100.0
	Total	153	100.0	

The results also revealed that 49.6% of the respondents agree and 50.4 disagree with the statement that "It is very difficult for the teacher to use alternative assessment at university level" (item 23; see Table 15). The mean for all the responses on this item was found to be 3.95, denoting a neither-agree-nor-disagree position. About 63% of the respondents agreed and 32% disagreed (at different levels) with the statement that "Alternative assessment methods seem less practical than traditional testing" (item 24; see Table 16). The mean for all the responses on this item was found to

Rezaee, Alavi, and Shabani

be 3.56 which can be considered as a weak inclination for all the respondents to agree with the statement. It was interesting to find that from the respondents, 83% agreed that "Using alternative assessment takes more time for the teacher in and out of the classroom than traditional testing" (item 27; see Table 17). Also a mean of 2.24 for all the responses on this item suggests a general agreement with the statement. This may be explained in relation to the responses on another item of the questionnaire. When asked to select the ideal number of students in a course in which they want to use alternative assessment approaches, (item 30), more than 58% selected less than 10 (see Table 18), perhaps indicating the demanding nature of alternative assessment (Bennett & Ba, 1996).

Table 15: Frequency of responses for item 23

		Frequency	Valid Percent	Cumulative Percent
Valid	1	14	9.2	9.2
	2	44	28.8	37.9
	3	18	11.8	49.7
	5	28	18.3	68.0
	6	35	22.9	90.8
	7	14	9.2	100.0
	Total	153	100.0	

Table 16: Frequency of responses for item 24

		Frequency	Valid Percent	Cumulative Percent
Valid	1	7	4.6	4.6
	2	62	40.5	45.1
	3	28	18.3	63.4
	4	7	4.6	68.0
	6	42	27.5	95.4
	7	7	4.6	100.0
	Total	153	100.0	

Alternative assessment or traditional testing

Table 17: Frequency of responses for item 27

		Frequency	Valid Percent	Cumulative Percent
Valid	1	61	39.9	39.9
	2	52	34.0	73.9
	3	14	9.2	83.0
	4	9	5.9	88.9
	5	7	4.6	93.5
	6	5	3.3	96.7
	7	5	3.3	100.0
	Total	153	100.0	

Table 18: Frequency of responses for item 30

	Frequency	Valid Percent	Cumulative Percent
Valid Less than 10	90	58.8	58.8
10-20	42	27.5	86.3
21-30	14	9.2	95.4
more than 50	7	4.6	100.0
Total	153	100.0	

None of the respondents strongly agreed with the statement "Results of traditional testing procedures demonstrate an objective picture of the students' progress in a given course" (item 19), and in fact only 7% of them agreed with the statement (Table 19). The mean for all the responses to this item was found to be 4.41, implying a mild disagreement with the statement. Of all the participants in the study, only 12% agreed with the statement "Traditional testing is an ethical and fair approach of evaluation" (item 6). Whereas 27% strongly disagreed and 27% disagreed with this statement, nobody strongly agreed with it (Table 20). The mean for all the responses on this item was found to be 5.45, indicating a general disagreement with the statement. However, 75% of the

Rezaee, Alavi, and Shabani

respondents agreed that "Alternative assessment is an ethical and fair approach of evaluation" (item 11). In addition, nobody strongly disagreed with the statement (Table 21). A mean of 2.80 for all the responses to this item suggests a general agreement with the statement.

Table 19: Frequency of responses for item 19

		Frequency	Valid Percent	Cumulative Percent
Valid	2	11	7.2	7.2
	3	36	23.5	30.7
	4	21	13.7	44.4
	5	57	37.3	81.7
	6	21	13.7	95.4
	7	7	4.6	100.0
	Total	153	100.0	

Table 20: Frequency of responses for item 6

		Frequency	Valid Percent	Cumulative Percent
Valid	2	4	2.6	2.6
	3	15	9.8	12.4
	4	15	9.8	22.2
	5	35	22.9	45.1
	6	42	27.5	72.5
	7	42	27.5	100.0
	Total	153	100.0	

Alternative assessment or traditional testing

Table 21: Frequency of responses for rtem 11

		Frequency	Valid Percent	Cumulative Percent
Valid	1	14	9.2	9.2
	2	77	50.3	59.5
	3	23	15.0	74.5
	4	18	11.8	86.3
	5	7	4.6	90.8
	6	14	9.2	100.0
	Total	153	100.0	

It was surprising to find that nobody agreed with the statement that "Ethics and ethical issues are adequately addressed in present approaches to language testing and evaluation" (item 12). In fact, only 22.9% of all the participants agreed with it (Table 22). A mean of 4.73 and a mode of 6 for all the responses on this item denote a general disagreement with the statement. Together with items 6 and 11, this finding implies the prevalence of traditional testing and a very less widespread application of alternative assessment even in the present classroom practice.

Table 22: Frequency of responses for item 12

		Frequency	Valid Percent	Cumulative Percent
Valid	2	7	4.6	4.6
	3	28	18.3	22.9
	4	36	23.5	46.4
	5	25	16.3	62.7
	6	43	28.1	90.8
	7	14	9.2	100.0
	Total	153	100.0	

Rezaee, Alavi, and Shabani

About 72.5% of the respondents agreed with the statement that "Since in traditional testing, students' performance and progress throughout the course will be assessed mainly in an end-of-the course exam (also known as final exam), and shown by a single score, traditional testing cannot be considered ethical and fair to students" (item 21), and only 4.6% strongly disagreed with it (Table 23). A mean of 2.9 and a mode of 1 for all the responses on this item denote a general agreement with the statement. Also, as can be seen in table 24, about 75% of the respondents agreed that despite the usefulness of alternative assessment to students, "In the end it is their scores that count in later decisions and judgments" (item 25). A mean of 2.27 for all the responses to this item denotes a general agreement with the statement.

Table 23: Frequency of responses for item 21

		Frequency	Valid Percent	Cumulative Percent
Valid	1	49	32.0	32.0
	2	22	14.4	46.4
	3	40	26.1	72.5
	4	7	4.6	77.1
	5	14	9.2	86.3
	6	14	9.2	95.4
	7	7	4.6	100.0
	Total	153	100.0	

Table 24: Frequency of responses for item 25

		Frequency	Valid Percent	Cumulative Percent
Valid	1	43	28.1	28.1
	2	71	46.4	74.5
	3	14	9.2	83.7
	4	11	7.2	90.8
	5	7	4.6	95.4
	6	7	4.6	100.0
	Total	153	100.0	

In fact, one prominent area of divergence between alternative assessment and traditional testing is the way the evaluation is reported. Whereas in traditional testing the emphasis is on the product (e.g., Culbertson & Jalongo, 1999), and evaluation is usually presented in the form of a single score, in alternative assessment, the assessment results are usually reported as a profile of the students' progress or process of learning. In terms of function, thereby, alternative assessment is formative rather than summative. They are at the same time less formal, and in terms of consequence, they are generally considered low-stakes (Alderson & Banerjee, 2001).

As indicated in Table 25, about 67% of the teachers agreed that "No matter how useful it may be to use alternative assessment, at the end of the course, teachers are expected to report scores for their students" (item 20). A mean of 2.84 for all the responses on this item indicates a general agreement with the statement. This may explain the prevalence of traditional testing even in the present classroom practices.

Table 25: Frequency of responses for item 20

		Frequency	Valid Percent	Cumulative Percent
Valid	1	36	23.5	23.5
	2	60	39.2	62.7
	3	7	4.6	67.3
	4	7	4.6	71.9
	5	28	18.3	90.2
	6	15	9.8	100.0
	Total	153	100.0	

Only 18% of the teachers expressed their disagreement with the statement that "Alternative assessment can be considered more as supplement rather than substitute for traditional testing" (item 22) (Table 26). The mean and mode for all the responses on this item were found to be 2.99 and 1, respectively, suggesting a general agreement with the statement. Also, it was found that 86% of the

Rezaee, Alavi, and Shabani

teachers agreed that "Alternative assessment and traditional testing are both necessary and needed, and both should be practiced in the classroom" (item 26) (Table 27). The mean and mode for all the participants' responses to this item were found to be 2.03 and 1, respectively, bespeaking a general agreement with the statement.

Table 26: Frequency of responses for item 22

		Frequency	Valid Percent	Cumulative Percent
Valid	1	39	25.5	25.5
	2	35	22.9	48.4
	3	30	19.6	68.0
	4	21	13.7	81.7
	5	7	4.6	86.3
	6	7	4.6	90.8
	7	14	9.2	100.0
	Total	153	100.0	

Table 27: Frequency of responses for item 26

		Frequency	Valid Percent	Cumulative Percent
Valid	1	70	45.8	45.8
	2	51	33.3	79.1
	3	11	7.2	86.3
	4	7	4.6	90.8
	5	7	4.6	95.4
	6	7	4.6	100.0
	Total	153	100.0	

The results suggest that, all in all, traditional testing still seems to be the more commonly practiced approach. Twenty-seven percent of the teachers *always* use it in their classes, and 36% checked that they *often* use it. None of the respondents selected *never* in response to how often they use traditional testing in their

classes (item 28). The mean for all the responses to this item was found to be 2.37, indicating a tendency for language teachers to *often* use some form of traditional testing in their classes. Relevant to this item is item 1 (i.e., "I am satisfied with the evaluation process and practices used in my classes"). In fact, nobody strongly disagreed that they are satisfied with the evaluation process and practices used in their classes, and about 60% of the teachers expressed their satisfaction with their classroom practice in this regard (Table 28). A mean of 3.41 for all the respondents, however, denotes a moderate feeling of satisfaction. Arguably, it may not be implausible to attribute this mildness of satisfaction to the findings on item 12.

Table 28: Frequency of responses for item 1

		Frequency	Valid Percent	Cumulative Percent
Valid	1	14	9.2	9.2
	2	35	22.9	32.0
	3	43	28.1	60.1
	4	25	16.3	76.5
	5	8	5.2	81.7
	6	28	18.3	100.0
	Total	153	100.0	

It was also found that only 13% of the teachers *always* use alternative assessment in their classes, and 35% of them checked that they *rarely* use it. The mean was found to be 2.85, denoting that teachers are inclined to *sometimes* use some form of alternative assessment in their classes (item 29).

About 67% of the teachers disagreed with the statement "All ethical standards are relative, to the degree that there are no permanent, universal, objective values and standards" (item 31). A total mean of 4.95 for all the responses indicates a general tendency to disagree with the statement. It was not surprising then to find that 74% of the respondents agreed that "There are at least some ethical

Rezaee, Alavi, and Shabani

values, standards, or principles that are not relative" (item 32). A mean of 2.67 for all the responses can be taken as a general tendency among all the participants to agree with the statement (see Tables 29 and 30 below).

Table 29: Frequency of responses for item 31

		Frequency	Valid Percent	Cumulative Percent
Valid	1	10	6.5	6.5
	2	30	19.6	26.1
	3	10	6.5	32.7
	5	12	7.8	40.5
	6	40	26.1	66.7
	7	51	33.3	100.0
	Total	153	100.0	

Table 30: Frequency of responses for item 32

		Frequency	Valid Percent	Cumulative Percent
Valid	1	50	32.7	32.7
	2	51	33.3	66.0
	3	12	7.8	73.9
	4	10	6.5	80.4
	6	30	19.6	100.0
	Total	153	100.0	

Considering the origin of ethics (item 33), about 30% believed that ethics comes from one's society, 14% believed that "God or religion" is the origin of ethics, 34% selected "rationality and logic", 16% checked "human nature designed by God", and 6% considered "human nature shaped by evolution" as the origin of ethics and ethical standards (see Table 31).

Alternative assessment or traditional testing

Table 31: Frequency of responses for item 33

		Frequency	Valid Percent	Cumulative Percent
Valid	one's society	51	30.3	33.3
	God or religion	21	13.8	47.1
	rationality and logic	51	33.3	80.4
	human nature designed by God	20	16.1	93.5
	human nature shaped by evolution	10	6.5	100.0
	Total	153	100.0	

In fact, this item on the origin of ethics was found to be one of the most challenging items in the questionnaire. It is worth noting at this point that in the earlier drafts of the questionnaire, there were more items directly addressing ethical issues and examining teachers' ethical orientation. Yet, the researchers later decided to remove some of these items as some respondents considered some questions personal, and expressed that they did not feel at ease with sharing what they considered their private ideas with others. Also, some respondents felt that some of the topics addressed in the questionnaire could not be answered directly. One of the teachers in the piloting stage commented, "I think such issues need quite a deal of contemplation, reasoning, discussion and exchange of ideas and viewpoints. The answers I provided are my own conceptions and prone to change on discussion."

Though not very popular with few respondents, item 33 was kept in the questionnaire, as it could provide useful insight into the interpretation of the results. One of the respondents believed that "ethical standards come from God, but not religion, for religions are mostly cultural affairs which eventually display disparity rather than unity. God alone places before us the options for ethical behavior and thinking. In this way ethics universalizes behavior and thought, far beyond the vagaries of human cultures . . . . Cultures are volatile expressions of how peoples are behaving collectively."

Another respondent commented that "As human being is a social-psychological creature, I believe, the instant (family-wide)

Rezaee, Alavi, and Shabani

and distant (city- and country-wide) environment can affect his perceptual and intellectual development which, in turn, shapes the values and standards he defines for himself. That is why I check both 'one's society and rationality and logic' which I assume as being closely interconnected."

In fact, the purpose of the study is not to prescribe any ethical standard to language testers or to get to any unanimous conclusion on the ethical policy to be pursued in the field. An investigation to the possible consistent interaction of theory and practice with regard to the stakeholders' understanding of ethics in language assessment can become the diving board, as it were, for further exploration into the place of ethics in the field which may ultimately lead to establishing more ethical assessment contexts. The complexity of this interface, as alarming as it may seem, should be considered as an awesome enterprise promising in different aspects. However, as in other areas of enquiry which deal with humans and their relations, the problem in the study of ethics and its place in language assessment lies in the definitions and basic denotations. In language assessment, high-stakes decisions are made and crucial choices are exercised. Decisions and choices bring responsibilities. People's beliefs result from their decisions in life, and making decisions implies the recognition of one's volition in making choices. Language testers therefore should take responsibility for holding certain beliefs (Deigh, 2010).

Responsibility, though, is best represented as a continuum. People are not equally responsible in certain situations with regard to the morality of their conduct. Generally, the responsibility is heavier on the shoulder of some of the stakeholders than the others. In effect, it is people's decisions based on their beliefs that bring responsibility for them. As a challenging area of enquiry immersed in ethical decisions, language assessment is a consequential process. The consequences are essentially built on decisions which in turn derive from beliefs. The crux of the matter is that prior to any ethical investigation into language assessment, the ethical beliefs of those involved should be subjected to scrutiny. Also, before expecting any change in people's behavior, measures must be taken to change their thoughts and this will not be possible unless there is a clear delineation of their thoughts. Ethical assessment will simply

remain a myth unless the stakeholders' conceptions of what ethical means and what ethical assessment is can get on the right track.

5. Conclusion

Thoughts can influence choices and choices bring forth responsibilities (Hughes, 2001). So, the choices made in relation to language assessment create responsibility, and this responsibility is indirectly influenced by the thoughts and ideological orientations of the chooser. Inasmuch as language assessment concerns conscious decisions on the part of stakeholders, the consequences brought about as a result of those decisions are, of necessity, attributable to the thoughts of the decision-makers. Therefore, this study was an attempt to examine what teachers think about alternative assessment and traditional testing and also how they generally perceive ethics and ethical standards.

Lynch (2001b) and Lynch and Shaw (2005) argue that traditional testing and alternative assessment follow two significantly different research paradigms. The former entails practices that follow the positivist perspective, whereas the later involves assessment practices following the interpretivist perspective. In fact, the distinction is represented in the terminology used in the related literature. The terms *test* and *assessment* have been used in the recent publications, particularly on validity and ethics, to denote two different perceptual orientations and research perspectives (Leung & Lewkowicz, 2006).

Alternative assessment requires and offers a different culture, an "assessment culture" rooted in an "epistemology of mind" (Lynch, 2001a, p. 229) which acknowledges the integration of teaching and assessment, and the active participation of students in developing their own assessment procedure and criteria. To McNamara (2001b), alternative assessment, "as normally understood" (p. 330), refers to "a movement, particularly in school contexts in the USA, away from the use of standardized multiple-choice tests in favor of more complex performance-based assessments" (p. 329). He then goes on to explain that to him the concept can go beyond this definition to include procedures which

"range from standardized tests to assessment activities without a measurement focus at all" (p. 330).

Generally, 90% of teachers use teacher-made tests, at least, once in a month (Rotham, 1995). Teachers use traditional testing to assign grades at the end of an instructional program. The purpose is to distinguish successful students from unsuccessful ones (Wilcox & Zielinski, 1997). In fact, the most obvious limitation of traditional testing is lack or paucity of student involvement in the process of testing. Lack of authenticity, limited focus to factual information which assesses skills in isolation, emphasizing lower-level comprehension, marginalization of thinking skills, decision-making competencies, attitudes and values, creativity and innovation, individual abilities, and social and interpersonal dimensions are the most prominent disadvantages of traditional testing (Atta-Alla, 2012).

Congruent with the findings of the study, despite the reported advantages of alternative assessment, an obstinate stigmatization and refusal of the traditional testing format may still seem a practice at odds with the common sense. In a comparative study of students' writing performance, Lee (2004) explicitly warns that for those students who show preference for writing on paper, "the traditional testing format still has to be kept" (p. 20). Also, it has been suggested that alternative assessment approaches do not necessarily guarantee equity for diverse populations (Darling-Hammond, 1994). Alternative assessment has nonetheless taken its fair share of criticism. Many parents still prefer to see their children's grades and question the elimination of grades (Culbertson & Jalongo, 1999). Also, As Hassel and Lourey (2005) contend, ". . . the pedagogical pursuit of alternative forms of assessment in elementary and secondary school may contribute to lack of preparedness for modes of evaluation at the college level (p. 4). The time-consuming procedure has also been frequently mentioned as one of the shortcomings of alternative assessment compared to paper-and-pencil tests (Kusimo, Ritter, Busick, Ferguson, Trumbull, & Solano-Flores, 2000). Another major criticisms leveled against alternative assessment is the inconsistency of marking criteria and grading schemes (Clapham, 2000). The wealth of information provided by alternative assessment approaches, however, is

undeniable, and as Hamayan (1995) believes, this information "must minimally serve as a context for a more valid interpretation of all standardized test results" (p. 212). This explains why alternative terminology has been used in the literature to refer to alternative assessment, such as supplementary or complementary. Alternative assessment procedures "do not take the place of summative assessment and standardized test scores, but complement these scores with a host of ongoing and comprehensive assessment measures (p. 2).

In fact, until a better proposal can be offered, alternative assessment and traditional testing can best be regarded as supplements rather than two ends of a continuum. Teachers can review students' portfolios and keep records of their reviews and then brief students accordingly on their progress (or the lack thereof). Based on their reviews, teachers can also brief administrators and parents on individual student growth. They can also design and use checklists and a set of descriptors to address the underlying points for which the portfolio has been used and keep record of students' progress (Tannenbaum, 1996).

References

Alderson, J. C., & Banerjee, J. (2001). Language testing and assessment (Part I). *Language Teaching, 34*, 213-236.

Alderson, J.C., & Wall, D. (1993). Does washback exist? *Applied Linguistics, 14*, 115-129.

Alderson, J.C., & Wall, D. (1996). [Special Issue]. Washback. *Language Testing, 13*(3), 280-297.

Atta-Alla, M. (2012). Investigating alternative assessment practices in preservice English teacher education programs. *International Journal of English Language Education, 1*(1), 68-81.

Bennett, D. & Ba, H. (1996). Student's voices on alternative assessment. *Connect: Supporting Student Participation, 98*, 14-16.

Bethlehem, J. (2008). *How accurate are self-selection web surveys?* (Discussion paper 08014). The Hague: Statistics Netherlands.

Brown, J. D., & Hudson, T. (1998). The alternatives in language assessment. *TESOL Quarterly, 32*(4), 653-675.

Rezaee, Alavi, and Shabani

Clapham, C. (2000). Assessment and testing. *Annual Review of Applied Linguistics, 20*, 147-161.

Culbertson, D. L., & Jalongo, R. M. (1999). But what's wrong with letter grades? *Childhood Education, 75*(3), 130-136.

Darling-Hammond, L. (1994). Performance-based assessment and educational equity. *Harvard Review, 64*, 5-30.

Davies, A. (1997). Demands of being professional in language testing. *Language Testing, 14*(3), 328-339.

Deigh, J. (2010). *An introduction to ethics*. Cambridge: Cambridge University Press.

Dörnyei, Z. (2003). *Questionnaires in second language Research: Construction, administration, and processing.* Mahwah, New Jersey: Lawrence Erlbaum Associates.

Elder, C. (1997). What does test bias have to do with fairness? *Language Testing, 14*(3), 261-277.

Hamayan, E. V. (1995). Approaches to alternative assessment. *Annual Review of Applied Linguistics, 15*, 212-226.

Hamp-Lyons, L. (1997). Washback, impact and validity: Ethical concerns. *Language Testing, 14*(3), 295-303.

Hancock, C. R. (1994). *Alternative assessment and second language study: What and why?* (Report No. ED376695 1994-07-00). Washington: Center for Applied Linguistics (CAL). (ERIC Clearinghouse on Languages and Linguistics Washington DC).

Hassel H. & Lourey, J. (2005). The Dea(r)th of student responsibility. *College Teaching, 53*(1), 2-14.

Holland, P. W., & Thayer, D. T. (1988). Differential item performance and the Mantel–Haenszel procedure. In H. Wainer, & H. I. Braun (Eds.), *Test validity* (pp. 129-145). Hillsdale, NJ: Erlbaum.

Holland, P. W., & Wainer, H. (Eds.) (1993). *Differential item functioning*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology, 24*, 417-441.

Hughes, G. J. (2001). *Aristotle: On ethics*. London: Routledge.

Kusimo, P., Ritter, M. G., Busick, K., Ferguson, C., Trumbull, E., & Solano-Flores, G. (2000). *Making assessment work for everyone: How to build on student strengths*. Austin: Assessment Laboratory Network Project of the Regional Educational Laboratories.

Lee, H. K. (2004). A comparative study of ESL writers' performance in a paper-based and a computer-delivered writing test. *Assessing Writing, 9*, 4-26.

Leung C., & Lewkowicz, J. (2006). Expanding horizons and unresolved conundrums: Language testing and assessment. *TESOL Quarterly, 40*(1), 211-234.

Lynch, B. K. (1997). In search of the ethical test. *Language Testing, 14*(3), 315-327.

Lynch, B. K. (2001a). The ethical potential of alternative assessment. In C. Elder, A. Brown, E. Grove, K. Hill, N. Iwashita, T. Lumley, T. McNamara, & K. O'Loughlin (Eds.), *Experimenting with uncertainty: Essays in honour of Alan Davies* (pp. 228-239). Cambridge: Cambridge University Press.

Lynch, B. K. (2001b). Rethinking assessment from a critical perspective. *Language Testing, 18*(4), 351-372.

Lynch, B. K., & Shaw, P. (2005). Portfolios, power, and ethics. *TESOL Quarterly, 39*(2), 263-297.

Mathew, R. (2004). Stakeholder involvement in language assessment: Does it improve ethicality? *Language Assessment Quarterly, 1*(2), 123-135.

McNamara, T. (2001a). Editorial: Rethinking alternative assessment. *Language Testing, 18*(4), 329-332.

McNamara, T. (2001b). Language assessment as social practice: Challenges for research. *Language Testing, 18(4)*, 333-349.

McNamara, T., & Roever, C. (2006). *Language assessment: The social dimension.* Malden, MA: Blackwell Publishing.

Messick, S. (1980). Test validity and the ethics of assessment. *American Psychologist, 35*, 1012-1027.

Messick, S. (1996). Validity and washback in language testing. *Language Testing, 13*(3), 241-256.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook.* London: Sage

Milne J., & Oberle K. (2005). Enhancing rigor in qualitative description: A case study. *Journal of Wound Ostomy Continence Nursing, 32*, 413-420.

Moss, P. A. (1992) Shifting conceptions of validity in educational measurement: implications for performance assessment. *Review of Educational Research, 62*, 229-258.

Neergaard, M., Olesen, F., Andersen, R., & Sondergaard, J. (2009). Qualitative description—The poor cousin of health research? *BMC Medical Research Methodology, 9*(1), 1-5.

Penfield, R. D., & Camilli, G. (2007). Differential item functioning and item bias. In C. R. Rao & S. Sinharay (Vol. Eds.), *Handbook of statistics*: Vol. 26 (pp. 125-167). Amsterdam: Elsevier.

Rezaee, Alavi, and Shabani

Rea-Dickins, P. (1997). So, why do we need relationships with stakeholders in language testing? A view from the UK. *Language Testing, 14*(3), 304-314.

Rotham, R. (1995). *Measuring up: Standards, assessment, and school reform*. San Francisco: Jossey-Bass.

Sandelowski, M. (2000). Whatever happened to qualitative description? *Research in Nursing and Health, 23,* 334-340.

Shohamy, E. (1997). Testing methods, testing consequences: Are they ethical? Are they fair? *Language Testing, 14*(3), 340-349.

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 use of language tests.* Harlow: Longman.

Stansfield, C. W. (1993). Ethics, standards, and professionalism in language testing. *Issues in Applied Linguistics, 4*(2), 15-30.

Tannenbaum, J. (1996). *Practical ideas on alternative assessment for ESL students. ERIC Digest.* ERIC Clearinghouse on Languages and Linguistics Washington DC. (No. ED395500). Retrieved July 5, 2012 from the World Wide Web: http://www.eric.ed.gov

Tsagari, D. (2004). *Is there life beyond language testing? An introduction to alternative language assessment.* CRILE Working Papers No. 58, Center for Research in Language Education (CRILE), Linguistics and English Language Department, Lancaster University.

van Daalen, M. (1999). Test Usefulness in Alternative Assessment. *Dialog on Language Instruction, 13*(1 & 2), 1-26.

Wall, D., & Alderson, J. C. (1993) Examining washback: The Sri Lankan impact study. *Language Testing, 10*(1), 41-69.

Watkins, M. W. (2000). *Monte Carlo PCA for parallel analysis (*version 2.5).   [Computer software]. State College, PA: Ed & Psych Associates.

Wilcox, S. K., & Zielinski, R. S. (1997). Using the assessment of students' learning to reshape teaching. *Mathematics Teacher*, *90*(3), 223-229.

Young, F.W. (1996). *ViSta "the visual statistics system".* [Computer software]. Retrieved July 22, 2013. Available from http://forrest.psych.unc.edu/research/index.html

Young, F. W., Ledesma, R., Molina, J. G., & Valero, P. (1997). ViSta: A visual statistics system. In R. A. Stine & J. Fox (Eds.), *Statistical computing environments for social research* (pp. 207-235). Thousand Oaks: Sage.

Appendix

Questionnaire

Dear EFL/ESL Teacher,

Thank you for accepting to complete this questionnaire. This questionnaire explores your ideas and opinions (as a teacher) on traditional testing and alternative assessment.

Part I

In items 1 to 27

1 means "strongly agree," and 7 means "completely disagree". If you choose 4, it denotes that you "neither agree nor disagree" with the statement.

1. I am satisfied with the evaluation process and practices used in my classes.
    Strongly Agree   ①   ②   ③   ④   ⑤   ⑥   ⑦   Strongly Disagree

2. Traditional testing is easy for correction.
    Strongly Agree   ①   ②   ③   ④   ⑤   ⑥   ⑦   Strongly Disagree

3. On the whole, traditional testing is beneficial to learners.
    Strongly Agree   ①   ②   ③   ④   ⑤   ⑥   ⑦   Strongly Disagree

4. Through traditional testing, teachers can provide students with ample feedback on their progress and performance throughout the course.
    Strongly Agree   ①   ②   ③   ④   ⑤   ⑥   ⑦   Strongly

Rezaee, Alavi, and Shabani

Disagree

5. Traditional testing is administrable in almost all situations and classes.

Strongly
Disagree

Strongly Agree   ①   ②   ③   ④   ⑤   ⑥   ⑦

6. Traditional testing is an ethical and fair approach of evaluation.

Strongly Agree   ①   ②   ③   ④   ⑤   ⑥   ⑦   Strongly
Disagree

7. Alternative assessment is easy for correction.

Strongly Agree   ①   ②   ③   ④   ⑤   ⑥   ⑦   Strongly
Disagree

8. On the whole, alternative assessment is beneficial to learners.

Strongly Agree   ①   ②   ③   ④   ⑤   ⑥   ⑦   Strongly
Disagree

9. Through alternative assessment, teachers can provide students with ample feedback on their progress and performance throughout the course.

Strongly Agree   ①   ②   ③   ④   ⑤   ⑥   ⑦   Strongly
Disagree

10. Alternative assessment is administrable in almost all situations and classes.

Strongly Agree   ①   ②   ③   ④   ⑤   ⑥   ⑦   Strongly
Disagree

11. Alternative assessment is an ethical and fair approach of evaluation.

Strongly Agree   ①   ②   ③   ④   ⑤   ⑥   ⑦   Strongly
Disagree

12. Ethics and ethical issues are adequately addressed in present approaches to language testing and evaluation.

Strongly Agree   ①   ②   ③   ④   ⑤   ⑥   ⑦   Strongly
Disagree

13. Exact results can be expected from traditional testing.

Alternative assessment or traditional testing

Strongly Agree   ①   ②   ③   ④   ⑤   ⑥   ⑦   Strongly Disagree

14. Exact results can be expected from alternative assessment.

Strongly Agree   ①   ②   ③   ④   ⑤   ⑥   ⑦   Strongly Disagree

15. Traditional testing adequately measures the learning outcomes.

Strongly Agree   ①   ②   ③   ④   ⑤   ⑥   ⑦   Strongly Disagree

16. Alternative assessment adequately measures the learning outcomes.

Strongly Agree   ①   ②   ③   ④   ⑤   ⑥   ⑦   Strongly Disagree

17. Alternative assessment forms an essential part of education for its flexibility and adjusting to the student learning styles and individual development.

Strongly Agree   ①   ②   ③   ④   ⑤   ⑥   ⑦   Strongly Disagree

18. Alternative assessment should be used in primary and secondary education (and not higher education, i.e. at university level).

Strongly Agree   ①   ②   ③   ④   ⑤   ⑥   ⑦   Strongly Disagree

19. Results of traditional testing procedures demonstrate an objective picture of the students' progress in a given course.

Strongly Agree   ①   ②   ③   ④   ⑤   ⑥   ⑦   Strongly Disagree

20. No matter how useful it may be to use alternative assessment methods, at the end of the course, teachers are expected to report scores for their students.

Strongly Agree   ①   ②   ③   ④   ⑤   ⑥   ⑦   Strongly Disagree

21. Since in traditional testing, students' performance and progress throughout the course will be assessed mainly in an end-of-the course exam (also known as final exam), and shown by a single score, traditional testing cannot be considered ethical and fair to students.

Rezaee, Alavi, and Shabani

Strongly Agree   ① ② ③ ④ ⑤ ⑥ ⑦   Strongly Disagree

22. Alternative assessment can be considered more as supplement to rather than substitutes for traditional testing.

Strongly Agree   ① ② ③ ④ ⑤ ⑥ ⑦   Strongly Disagree

23. It is very difficult for the teacher to use alternative assessment at university level.

Strongly Agree   ① ② ③ ④ ⑤ ⑥ ⑦   Strongly Disagree

24. Alternative assessment methods seem less practical than traditional testing.

Strongly Agree   ① ② ③ ④ ⑤ ⑥ ⑦   Strongly Disagree

25. No matter how useful alternative assessment can be to students' learning, in the end it is their scores that count in later decisions and judgments.

Strongly Agree   ① ② ③ ④ ⑤ ⑥ ⑦   Strongly Disagree

26. Alternative assessment and traditional testing strategies are both necessary and needed, and both should be practiced in the classroom.

Strongly Agree   ① ② ③ ④ ⑤ ⑥ ⑦   Strongly Disagree

27. Using alternative assessment takes more time for the teacher in and out of the classroom than traditional testing.

Strongly Agree   ① ② ③ ④ ⑤ ⑥ ⑦   Strongly Disagree

Part II

28. How often do you use some form of traditional testing in your classes?

Always ① ② ③ ④ ⑤ Never

29. How often do you use some form of alternative assessment in your classes?
Always ① ② ③ ④ ⑤ Never

30. Think of a course in which you want to use alternative assessment approaches. Ideally, how many students will be in this course?

☐ Less than 10
☐ 10-20
☐ 21-30
☐ 31-50
☐ More than 50

Part III

31. All ethical standards are relative, to the degree that there are no permanent, universal, objective values and standards.
Strongly Agree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Disagree

32. There are at least some ethical values, standards or principles that are not relative.
Strongly Agree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Disagree

33. In your opinion, where do ethical standards come from?

☐ One's society
☐ God or religion
☐ Rationality and logic
☐ Human nature that has been designed by God
☐ Human nature that has been shaped by millions of years of evolution

Rezaee, Alavi, and Shabani

Part IV

Gender
☐ Male   ☐ Female

Age
☐ Below 20   ☐ 20-29   ☐ 30-39   ☐ 40-49   ☐ 50-59
          ☐ 60-69   ☐ Over 70

Years of teaching experience
☐ 1-5   ☐ 6-11   ☐ 11+

Major (last degree)
☐ TESL/TEFL   ☐ Linguistics   ☐ Literature   ☐ Translation
☐ Other: _____

Last degree obtained
☐ B.A.   ☐ M.A.   ☐ Ph.D.   ☐ Other:
_____