

Iranian TOEFL iBT and the IELTS Teachers' Views on the Structure of the TOEFL iBT and IELTS Receptive and Productive Sections in terms of Dynamic and Static Assessment

Arezoo Daneshvar¹, Mohammad Sadegh Bagheri^{2*}, Firooz Sadighi³,
Lotfollah Yarmohammadi⁴, Mortaza Yamini⁵

¹ Ph.D. candidate, Department of English Language, Shiraz Branch, Islamic Azad University, Shiraz, Iran

² Assistant Professor of TEFL, Department of English Language, Shiraz Branch, Islamic Azad University, Shiraz, Iran

³ Professor of Applied Linguistics, Department of English Language, Shiraz Branch, Islamic Azad University, Shiraz, Iran

⁴ Professor of Applied Linguistics, Department of English Language, Shiraz Branch, Islamic Azad University, Shiraz, Iran

⁵ Assistant Professor of TEFL, Department of English Language, Zand Institute of Higher Education, Shiraz, Iran

Received: 15 March, 2020

Accepted: 20 July, 2020

Abstract

This mixed methods design study investigated Iranian TOEFL iBT and IELTS teachers' views on the structure of the TOEFL iBT and IELTS receptive and productive sections from the yardsticks of dynamic and static assessment. It also examined the conformity level of the receptive and productive sections of TOEFL iBT and IELTS to dynamic assessment and static assessment standards. To achieve the objectives of the study, 100 information-rich TOEFL iBT and IELTS teachers selected based on purposive and snowball sampling completed two 8-item researcher-made questionnaires on the underlying features of these exams. To cross-validate the quantitative results, we performed semi-structured interviews with 10 informed teachers selected through purposive sampling from among the questionnaire respondents. The semi-structured data were content analyzed using a researcher-made framework categorizing the distinctive dynamic assessment and static assessment features. The results of the qualitative phase corroborated those of the quantitative part revealing that these exams mainly conform to static assessment tenets and that they enjoy only a few dynamic assessment features. The pedagogical implications of the findings are also explicated.

Keywords: Dynamic assessment, IELTS teachers, Productive skills, Receptive skills, Static assessment TOEFL iBT

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

INTRODUCTION

Dynamic assessment (henceforth, DA) dawned as a complement to static assessment (Minnaert, 2002). Notwithstanding wide recognition of the effectiveness of dynamic assessment in revealing learners' current ability and learnability, no research studies have thus far been conducted on the design and administration of the TOEFL iBT and the IELTS as large-scale standardized tests in reference to the DA tenets. Hence, this study intended to fill in the missing gaps in the literature by investigating the degree of conformity of the TOEFL iBT and IELTS receptive and productive sections to SA and DA features. More specifically, this study sought to provide answers to our guiding research questions as follows:

- RQ1: Do the EFL teachers perceive that the receptive and productive modes of the TOEFL iBT observe the DA criteria?
- RQ2: Do the EFL teachers perceive that the receptive and productive modes of the IELTS observe the DA criteria?
- RQ3: Do the EFL teachers agree on the observation of SA criteria in the design of the receptive and productive modes of the TOEFL iBT?
- RQ4: Do the EFL teachers agree on the observation of SA criteria in the design of the receptive and productive modes of the IELTS ?
- RQ5: Is there any statistically significant difference between these two sections of the TOEFL iBT and the IELTS as to their compliance with DA and SA standards?

The findings of the present study rebound to IELTS and TOEFL preparation programs and training courses. Likewise, the international language assessment organizations in charge of the TOEFL iBT and IELTS administration may find the results of this study useful in perusing new avenues to incorporate DA tenets into these exams.

Ebadi and Rahimi (2019) examined the short and long-term effects of online dynamic assessment on EFL learners' IELTS academic writing skills as well as their beliefs concerning the influence of online synchronous DA on academic writing skills employing a sequential exploratory mixed-methods approach. The study results revealed a significant improvement in task achievement, cohesion and coherence, lexicon, and grammatical range and accuracy of both writing task 1 and task 2. Moreover, the results indicated that the language learners were strongly supportive of the positive effects of online DA on academic writing skills.

Additionally, Khoshsima and Izadi (2014) investigated the DA and SA influence on EFL learners' listening comprehension. The researchers realized the first two DA groups significantly outperformed the one assessed by static assessment and that predicting learners' learnability was feasible through dynamic assessment.

In another line of research, Ajideh and Nourdad (2012) investigated the effects of dynamic assessment on Iranian university students' reading comprehension. They concluded that dynamic assessment improved the language learners' reading comprehension regardless of their proficiency levels immediately and over time.

In 2002, Kozulin and Garb researched the effect of testers' mediation on the testees' performance in a standardized pre-test without any help and an alternative post-test, with some guidance given by testers subsequently. The findings showed that students had better performance in the post-test than the pre-test, implying that the students with average pre-test and modest gain scores should be taught 'learning how to learn' strategies.

As evident in the studies reported above, no empirical studies have investigated the DA and SA properties of high-stakes tests and the feasibility of reforming them based on DA standards to manifest test-takers' present and future abilities.

METHODS

A mixed methods design was employed to answer the research questions of this study. The study consisted of two quantitative and qualitative phases. In the quantitative phase, 100 qualified TOEFL iBT, and IELTS teachers completed two researcher-made questionnaires. In the second phase of the study, the qualitative data were collected from 10 IELTS and TOEFL iBT teachers answering 13 semi-structured interview questions to the saturation point.

Participants

The quantitative study sample consisted of 100 TOEFL iBT and IELTS teachers, some of whom were IELTS examiners and invigilators, based on Schwab's (1980) recommendation insisting on a 1 to 10 item-to-response ratio for each set of scales to be factor analyzed. They were carefully chosen through purposive and snowball sampling in some cases. Of 100 teachers, 97 were teaching in renowned language institutes in Iran, and three of them were Iranian TOEFL iBT and IELTS teachers living and working in Canada and America. They were 68 male and 32 female teachers in the 30-50 age range with 6-10 years of teaching experience in teaching the TOEFL iBT and IELTS. Besides, the majority of the participants held MA in English teaching and the rest of them were either B.A or Ph.D. holders.

The qualitative phase participants were 10 TOEFL iBT and IELTS teachers (five males, and five females) with the age range of 30-50 selected through purposive sampling from among the questionnaire respondents having 6 to 10 years of teaching experience. Likewise, two of them held MA in English teaching and the rest of them were either Ph.D. candidates or Ph.D. holders in this field. (Schwab, 1980)

Instruments

The data collection instruments were two 8-item questionnaires written in English to extract teachers' views concerning dynamic and static features of the receptive and productive

sections of the TOEFL iBT and the IELTS. Furthermore, semi-structured interviews were conducted to cross-validate the quantitative results.

Questionnaires

The first study instruments were two eight-item questionnaires consisting of two parts constructed by the researchers in the study with the reliability rates of 0.78 and 0.75, respectively. The questionnaires were constructed based on the reviewed literature on dynamic and static assessments and the researcher-compiled framework following some interviews done with three university professors of applied linguistics expert in dynamic and static assessments. The content validity of the questionnaires was checked by asking three professors of applied linguistics to comment on the items. Their input and suggestions were taken into account. It is worth mentioning that before the main study, a pilot study was done with a sample of 10 IELTS and TOEFL iBT teachers, examiners, and raters who were asked to complete two 10-item questionnaires so as to examine the reliability of the questionnaires. Given the responses provided, two items of each questionnaire which had either negative or low correlations and were loaded under a wrong factor in the factor analysis results were removed. The 8-item questionnaires whose content validity was confirmed by three experts in the field of applied linguistics were then utilized. The first part of the questionnaire captured the participants' demographic characteristics. The second part entailed eight 5-point Likert scale items on their views concerning the DA and SA traits of the TOEFL iBT and the IELTS receptive and productive sections.

Semi-structured Interviews

The second instrument used in this study was the semi-structured interview. For the purposes of this study, 10 one-session face-to-face interviews were conducted in English at interviewees' offices each for about 45 minutes. Interviews were performed with 10 TOEFL iBT and IELTS teachers based on the guide-

lines of a successful interview provided by Bingham and Moore (1959). (Bingham et al., 1959) Interviewees were asked 13 questions the validity of which was approved by four university professors of applied linguistics. The questions were designed based on the classifications of dynamic and static assessments in the researcher-compiled framework which intended to explore their views regarding the TOEFL iBT and IELTS exams' formats and structure as well as the extent of the conformity of their receptive and productive sections to static and dynamic assessment principles.

The Qualitative Framework

There was no comprehensive framework available to scrutinize the TOEFL iBT and the IELTS based on the DA and SA principles. Hence, the researchers designed and employed a framework based on the distinctive features of DA and SA put forth by sophisticated researchers and experts to serve the purpose.

Data Analysis

The data analysis was conducted employing the 24th version of the Statistical Package for the Social Sciences (SPSS) in the quantitative stage. Independent sample *t*-tests were run to examine the means of the receptive sections of both exams as well as those of the productive sections to compare their conformity to SA and DA features. Moreover, the contents of interviews on the DA and SA features of the TOEFL iBT and the IELTS were qualitatively scrutinized both by the main researcher and two of the researchers in the study, and a well-

informed auditor utilizing a researcher-devised framework categorizing SA and DA features.

RESULTS

Results of the First, Second, Third and Fourth Research Questions

All the interviews were transcribed and the interview data were content analyzed.

Data-Collection Procedures

The data for the quantitative phase was collected through the questionnaires. The participants in the quantitative phase answered either interactively designed questionnaires sent to them through e-mails or paper and pencil questionnaires completely and meticulously. It took them approximately three minutes to fill out the questionnaires. The data collection lasted over a period of three months from November 2018 to February 2019.

Concerning semi-structured interviews, 10 TOEFL iBT and IELTS teachers were selected from among the respondents to the questionnaire of the study. To ensure the anonymity of the participants, we employed the codes of T1, T2T10 to quote the transcripts. Moreover, the respondents' consent was sought to audio-record the interviews. Explaining the purpose of the interview, the researcher (interviewer) assured the participants that their responses would be kept confidential.

Quantitative Results of The TOEFL iBT Questionnaire

Table 1 shows general descriptive data of the TOEFL iBT questionnaire. As items 7 and 10 in the TOEFL iBT section had a negative correlation, they were deleted

Table 1.
Descriptive Statistics of the TOEFL iBT Questionnaire

| | TOEFL Item 1 | TOEFL Item 2 | TOEFL Item 3 | TOEFL Item 4 | TOEFL Item 5 | TOEFL Item 6 | TOEFL Item 8 | TOEFL Item 9 |
|---------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| N Valid | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| Missing | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Mean | 3.8000 | 4.0300 | 3.7400 | 4.5700 | 3.7500 | 4.5700 | 3.3200 | 4.4300 |
| Std. Deviation | 1.11916 | .92611 | .71943 | .55514 | .97830 | .59041 | 1.08134 | .70000 |

Descriptive Statistics of the Receptive Sections

Table 2 depicts the descriptive statistics of the TOEFL iBT questionnaire items related to receptive sections.

Table 2.

Descriptive Statistics of the TOEFL Receptive Sections

| Items | Strongly Disagree % | Disagree % | Undecided % | Agree % | St |
|-------|---------------------|------------|-------------|---------|----|
| 1 | 1 | 18 | 14 | 34 | |
| 2 | 0 | 8 | 17 | 39 | |
| 4 | 0 | 0 | 3 | 37 | |
| 6 | 0 | 0 | 5 | 33 | |

The first item of this questionnaire states that the TOEFL iBT receptive sections mainly focus on a comparison of learners' performance with their peers' performance. As illustrated in Table 2, this static feature was stressed by 67% of the teachers.

The second question of the questionnaire mentions that the TOEFL iBT receptive sections zoom in on predicting testees' future success without any *graduated* hints to check their performance improvement. According to Table 3 above, 75% of the participants agreed on this static feature of this exam.

The fourth questionnaire item accentuates the objective result interpretation

in the TOEFL iBT receptive sections. As it is evident, an overwhelming majority of teachers (97%) held this static feature in high regard.

The sixth item of the TOEFL iBT questionnaire highlights markers' neutral stance in the receptive sections. The great majority of the teachers (95%) agreed on this SA feature of these sections.

Descriptive Statistics of the Productive Sections

Table 3 depicts the descriptive statistics of the TOEFL iBT questionnaire items related to productive sections.

Table 3.

Descriptive Statistics of the TOEFL Productive Sections

| Items | Strongly Disagree % | Disagree % | Undecided % | Agree % | St |
|-------|---------------------|------------|-------------|---------|----|
| 3 | 1 | 8 | 12 | 74 | |
| 5 | 0 | 11 | 30 | 32 | |

Item 3 of the questionnaire implies that the focus of the productive sections of the TOEFL iBT is on predicting testees' potential success through some hints. As seen in Table 4, 79% of the participants concurred on this DA feature.

The fifth questionnaire item states that the TOEFL iBT productive sections are subjectively assessed. Following the descriptive

statistics presented in Table 4, 59% of all the teachers accorded high priority on this DA feature.

Descriptive Statistics of the TOEFL iBT Entire Exam

Table 4 depicts the descriptive statistics of the entire TOEFL iBT concerning items 8 and 9.

Table 4.

Descriptive Statistics of the TOEFL Entire Exam

| Item | Strongly Disagree % | Disagree % | Undecided % | Agree % | St |
|------|---------------------|------------|-------------|---------|----|
| 8 | 7 | 19 | 16 | 51 | |
| 9 | 1 | 2 | 0 | 47 | |

The eighth item of the questionnaire highlights the evaluation of each test-taker compared to himself. Evidently, about half of the participants perceive this point as a SA feature.

Item 9 of the questionnaire focuses on the standardized examining process in all sections of the TOEFL iBT except its speaking section.

Here again, as shown in Table 4, 97% of all

participants were favorably disposed to this static characteristic of the TOEFL iBT exam.

Quantitative Results of the IELTS Questionnaire

Table 5 demonstrates the descriptive data of the IELTS questionnaire. Given the low correlation of items 5 and 10, they were deleted.

Table 5.

Descriptive Statistics of the IELTS Questionnaire

| | IELTS Item 1 | IELTS Item 2 | IELTS Item 3 | IELTS Item 4 | IELTS Item 6 | IELTS Item 7 | IELTS Item 8 | IELTS Item 9 |
|---------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| N Valid | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| Missing | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Mean | 3.5000 | 4.1600 | 3.7700 | 4.5300 | 4.4300 | 3.7000 | 3.1600 | 4.3800 |
| Std. Deviation | .91563 | .88443 | .66447 | .59382 | .71428 | .70353 | 1.08916 | .82609 |

Table 6 exhibits the descriptive statistics of the questionnaire items related to the IELTS receptive sections.

Table 6.

Descriptive Statistics of the IELTS Receptive Sections

| Item | Strongly Disagree % | Disagree % | Undecided % | Agree % | § |
|------|---------------------|------------|-------------|---------|---|
| 1 | 3 | 17 | 10 | 67 | 3 |
| 2 | 0 | 7 | 11 | 41 | 4 |
| 4 | 0 | 1 | 2 | 40 | 5 |
| 6 | 0 | 3 | 4 | 40 | 5 |

comparing learners' performance with their peers' capabilities. According to Table 6, 70% of the participants agreed on this item representing SA characteristics.

The second item of the questionnaire claims that the IELTS receptive sections aim at predicting testees' future success without any intervention. Most of the participants (82%) unanimously approved of this SA feature of the IELTS.

Item 4 of the questionnaire stresses the objective interpretation of results in the IELTS receptive sections. As it is evident, the great

majority of the participants (97%) concurred on this point.

Item 6 of the questionnaire stresses the neutrality of the marker in the IELTS receptive sections. The results displayed in Table 6 revealed that the majority of the participants (93%) highlighted this SA feature of the IELTS.

Descriptive Statistics of the IELTS Productive Sections

Table 7 shows the descriptive statistics of the questionnaire items related to the IELTS productive sections.

Table 7.**Descriptive Statistics of IELTS Productive Sections**

| Item | Strongly Disagree % | Disagree % | Undecided % | Agree % | St |
|------|---------------------|------------|-------------|---------|----|
| 3 | 1 | 8 | 6 | 83 | |
| 7 | 0 | 10 | 14 | 72 | |

The third questionnaire item states that the IELTS productive sections mainly focus on predicting testees' potential success by offering some guidance. As shown in Table 7, 85% of the participants confirmed this DA feature.

Item 7 of the IELTS questionnaire accentuates examiners' interactive role in the IELTS productive sections. Following the descriptive

results, over three-quarters of the participants showed agreement on this point.

Descriptive Statistics of the IELTS Entire Exam

Table 8 depicts the descriptive statistics of the entire IELTS concerning items 8 and 9.

Table 8.**Descriptive Statistics of the IELTS Entire Exam**

| Item | Strongly Disagree % | Disagree % | Undecided % | Agree % | S |
|------|---------------------|------------|-------------|---------|---|
| 8 | 11 | 18 | 16 | 54 | 1 |
| 9 | 1 | 4 | 4 | 38 | 5 |

Item 8 of the questionnaire stresses that in the IELTS; each test-taker is evaluated against himself. Based on the results of the table above, about half of the participants (55%) agreed on this item, which was considered to represent a SA feature because it was reversed.

The ninth questionnaire item emphasizes that the examining process is the same in the IELTS reading, listening, and writing sections. As Table 8 depicts, a higher proportion of respondents (91%) concurred on this SA feature of the IELTS.

Results of the Fifth Research Question

Two independent sample *t*-tests were run to find the answer to this research question. To compare the means of the receptive sections of both exams, an independent sample *t*-test was conducted. As seen in Table 9, the TOEFL iBT receptive section enjoyed a higher mean score than that of the IELTS.

Table 9.**TOEFL iBT and the IELTS Receptive Sections' Statistics**

| | Test type | N | Mean | Std. Deviation | Std. Error Mean |
|-----------|-------------|-----|---------|----------------|-----------------|
| Receptive | 1.00(TOEFL) | 100 | 16.9700 | 2.49628 | .24963 |
| | 2.00(IELTS) | 100 | 16.6200 | 2.17321 | .21732 |

Table 10.

Receptive Sections' Independent Samples t-test

| Levene's Test for Equality of Variances | | | | | t-test for Equality of Means | | | | | |
|---|---------------------------------------|-------|------|-------|------------------------------|-----------------------|--------------------------|---|---------|---------|
| | F | Sig. | T | df | Sig. (2-tailed) | Mean dif- ferences | Std. Error Difference | 95% Confidence Interval of the Difference | | |
| | | | | | | | | Lower | Upper | |
| Receptive | Equal variance assumed | 4.121 | .044 | 1.057 | 198 | .292 | .35000 | 33097 | -.30268 | 1.00268 |
| | Equal variance not as- sumed | | | 1.057 | 194.315 | .292 | .35000 | 33097 | -.30276 | 1.00276 |

As viewed in Table 10, there was an insignificant difference in the results of the concurrence of the TOEFL iBT (M=16.9, SD= 2.4) and IELTS receptive sections (M= 16.6, SD= 2.1); $t(198) = 1.05$, $p = .29$ with SA and DA tenets. The results imply that the Iranian IELTS and TOEFL teachers were of the same opinion about the observation of SA standards

in the receptive sections of both the TOEFL iBT and IELTS exams.

Furthermore, to compare the two means of the TOEFL iBT and IELTS productive sections, an independent t-test was conducted. Table 12 illustrates the mean scores of the productive sections of the IELTS and the TOEFL iBT exam.

Table 11.

TOEFL iBT and IELTS Productive Sections' Statistics

| | Test type | N | Mean | Std. Deviation | Std. Error Mean |
|------------|--------------|-----|--------|----------------|-----------------|
| Productive | 1.00 (TOEFL) | 100 | 7.4900 | 1.35955 | .13596 |
| | 2.00 (IELTS) | 100 | 7.4700 | 1.12326 | .11233 |

As seen in Table 11, the TOEFL iBT productive sections had a slightly higher mean

score than the productive sections.

Table 12.

Productive Sections' Independent Samples t-test

| Levene's Test for Equality of Variances | | | | | t-test for Equality of Means | | | | | |
|---|---------------------------------------|-------|------|------|------------------------------|-----------------------|--------------------------|---|---------|--------|
| | F | Sig. | t | Df | Sig. (2-tailed) | Mean dif- ferences | Std. Error Difference | 95% Confidence Interval of the Difference | | |
| | | | | | | | | Lower | Upper | |
| Productive | Equal variance assumed | 6.294 | .013 | .113 | 198 | .910 | .02000 | .17635 | -.32777 | .36777 |
| | Equal variance not as- sumed | | | .113 | 191.197 | .910 | .02000 | .17635 | -.32785 | .36785 |

As can be seen in Table 12, the results of the t-test revealed an insignificant difference in the results of compliance of the productive sections of the TOEFL iBT (M=7.4, SD= 1.3) and IELTS (M=7.4, SD=1.1); $t(198) = .11, p = .91$ with SA and DA tenets. The findings implicitly suggest that the participants unanimously confirmed that there was no discrepancy in the conformity of the productive sections of the TOEFL iBT and IELTS exams to SA and DA principles.

Results of the TOEFL iBT and the IELTS Semi-structured Interviews

To cross-validate the quantitative results of the TOEFL iBT and the IELTS receptive and productive sections, semi-structured interviews were conducted. The interviewees were asked 13 questions (see Appendix C) the findings of which overwhelmingly substantiated those of the quantitative stage. As demonstrated in Table 14, there was substantial unanimity among interviewees upon the observation of a majority of static features in the TOEFL iBT and the IELTS receptive sections as well as a few DA features in the productive sections of these two exams. Nevertheless, they did not unanimously agree that these two exams compare test-takers with their peers, and consider testees' non-intellective factors.

In simpler terms, the first interview question was whether these two exams assess test-takers' ZPD (testees' learnability) or ZOA (current ability). All of the participants emphasized that testees' current ability is merely assessed in these exams, which is one of SA standards highlighted by Vygotsky (1998). The following is a sample excerpt from one of the interviewees.

T2: *In my opinion, current ability...maybe the third part of the IELTS interview is somehow interactive or maybe umm... in the integrated sections of the TOEFL iBT, the purpose is evaluating testees' learning potential but it doesn't give them any feedback and what's more, it is computerized so...there's no real interaction...*

The second interview question asked whether non-intellective factors such as learners' socio-economic level, intrinsic motivation, emotions such as anxiety, frustration, personality, need for mastery, individuals' background, teaching style, etc. are considered in the TOEFL iBT and the IELTS. Generally, it was thought that these factors which were indicative of a DA trait stated by (D Tzuriel, 2001) were somewhat taken into account in the design of these exams. The following excerpt further illustrates the point.

T3: *to some extent, personal traits and non-intellective factors of testees are considered in the test designs of these exams ...you know...some testees are turtle-like ...but they answer questions more carefully while some others are horse type...they answer quickly but they make more mistakes...they consider these points ...or even the fact that some are introverts but others are extrovert...what is their learning style...and many other points that's why there are different types of questions in each section...*

The third question asked whether testees are assessed in relation to others or themselves in these exams. The interviewees had different views in this regard. The following example can help better clarify the point.

T5: *Let me see....well in the IELTS ...yes testees are compared with othersmmm... the same applies to the TOEFL iBT but well.....maybe in the integrated sections testees' performance can be assessed with their own independent performance...but you know in the TOEFL iBT the testees' percentile is not offered as in PBT so I'm not sure they are compared with their peers...uh ...I don't know..*

The fourth question dealt with what is emphasized in these exams. All interviewees believed that these two exams place emphasis on the end-product. This is, in fact, one of the SA characteristics mentioned by (D. Tzuriel et al.,

1987). One of the teacher's accounts regarding the focal points of the two exams in questions is as follows:

T7: Both exams emphasize end-products but maybe in the TOEFL iBT integrated sections that students are given some input such as vocabulary, grammar or even information in the reading and listening passages, their learning process is considered, too.

The fifth question targeted the testing context of these exams. All interviewees underlined their static and standardized context except for the third part of the IELTS interview, which is communicative and slightly dynamic. As an example, one interviewee stated:

T10: To me, IELTS is more authentic and real-life but both exams are standard and formal...But I think the IELTS speaking section especially its third part is more interactive.

The sixth interview question addressed the way test-takers' performances are assessed. The general belief was that the performance evaluation is done based on particular norms in these two exams, which is in line with SA tenets underlined by (Morrison, 2001) One interviewee's quote is as follows:

T2: All sections of these exams are assessed based on special rubrics and norms so they can classify test-takers.

The seventh question dealt with testees' performance interpretation. The interviewees maintained that despite mostly objective interpretation of testees' performance in these exams, the subjective analysis could be somewhat attributed to their productive sections conforming to one of the DA criteria asserted by Tzuril (2001). One teacher commented that:

T3: TOEFL and IELTS objectively evaluate receptive skills of testees but the evaluation of productive sections can be a little bit subjective in spite of the criteria released for each exam...especially

in the IELTS interview which is face-to-face...

The eighth question inspected the examiner's stance in these exams. All interviewees stressed the TOEFL iBT examiners' neutrality since the oral assessment is done based on the interview record by examiners in non-authentic contexts in line with SA principles. As for the IELTS, most interviewees accentuated the examiners' neutrality, but some referred to the interviewers' partial interaction with test-takers in the third part of the interview. Inasmuch as the tester-testee interaction has been accentuated as a DA property by Morrison (2001), this can prove that the third part of the IELTS interview is slightly in accordance with DA standards. One of the interviewee's perspectives concerning this item is reported in the following interview extract.

T6: In TOEFL iBT, the examiner is robot-like and neutral....in IELTS, the examiner is a human so he cannot be neutral especially in the third part of the IELTS interview that is based on interaction and discussion...I mean in general it's neutral in IELTS too but not in the third part.

In the ninth question, the interviewees were asked whether the tasks in the IELTS and TOEFL iBT exams were constructed based on psychometric properties, i.e., reliability and validity, a SA trait stated by (Grigorenko et al., 1998), or based on testees' learnability from their mistakes. There was a unanimous agreement among the interviewees that only psychometric properties were considered while designing these exams. As an example, one teacher of the opinion that:

T7: TOEFL iBT tasks are based on psychometric properties such as reliability, validity and even it has psychometricians and statistics experts that is one of the strengths of this exam. The same applies to the IELTS.

The tenth question targeted the focus of these high-stakes exams. In this regard, most of interviewees stated that these exams do not merely focus on measuring testees' isolated competencies i.e. listening, reading, writing and speaking skills. Rather, they concentrate on test-takers' whole features including cognitive and non-intellective perceived as one of the DA characteristics by Morrison (2001)... For instance, one interviewee asserted that:

T10: *I thinkisolated competencies are their focus but in TOEFL iBT integrated sections the focus is mostly on the whole learner since ...mmm...the exam seeks to check whether the testee has got the points mentioned in the reading and listening passages or not...you know it checks their cognitive ability...*

The eleventh interview question concerned the nature of assessment in these exams. That is, whether assessment is affected by examiners' intervention, and test-takers- examiners' interaction, or examiners' predetermined questions without intervention. All interviewees claimed that testees are only evaluated by means of some fixed questions with no real interaction. This is one of the SA characteristics stated by Sternberg and (Grigorenko et al., 1998), in the TOEFL iBT. Although the interviewees believed that the same applies to the IELTS, they stated the third part of the IELTS interview is partially assessed through examiner-examinee face-to-face interactions. As mentioned before, the tester-testee interaction is one of the DA properties mentioned by (Haywood et al., 2006). Concerning the examiner-examinee interaction, one of the teachers maintained that:

T1: *Surely, mere assessment... through fixed questions because questions are not tuned based on individuals' needs and the time is fixed too especially in TOEFL iBT that the voice is not recorded after the given time....*

Question twelve asked whether TOEFL iBT and IELTS test-takers are given any feedback, or they are only informed about their overall score in relation to others. Interviewees

believed that the TOEFL iBT test-takers only get feedback on their productive performance besides their general score after the exam rather than during the assessment. This is one of the SA traits put forth by Haywood and Lidz (2006). Moreover, they thought the IELTS only provides test-takers with a general score showing test-takers' status in relation to others. In point of fact, this is one of the SA characteristics mentioned by (Merrell et al., 2012). The following interview excerpt pertains to how feedback is viewed in these two exams:

T1: *As far as I'm concerned, general score...but there is a part in the TOEFL grade report classifying testees as novice, expert, but it doesn't report what grammatical and lexical mistakes they made of course...*

The last question dealt with what is assessed in these exams. All interviewees believed that test-takers' independent performance is assessed in both exams in line with SA principles stated by (Vygotsky, 1998). However, some claimed that the TOEFL iBT integrated sections can be modified by offering *graduated prompts* based on DA principles to evaluate test-takers' assisted performance in relation to their independent performance. On this point, one teacher stated that:

T4: *Independent performance ...uhhh...but the TOEFL iBT integrated parts and IELTS interviews ...uh...especially part 3 assess interactive skills ...I think.*

It is noteworthy that the interview results were in line with those of quantitative ones highlighting the dominance of the conformity of the IELTS and TOEFL iBT receptive sections to SA tenets. Moreover, the interview findings were to some extent consistent with the ones obtained in the quantitative phase of the study indicating the interviewees' emphasis on the observation of only a few DA features in the productive sections of these two exams.

The major SA and DA features of the TOEFL iBT and the IELTS from the inter-

viewees' perspectives have been summarized in Table 13.

Table 13.
Teachers' Main Views about the TOEFL iBT and the IELTS

| No. | Category | TOEFL iBT Conformity to | | TOEFL iBT Percent- age of agreement | IELTS Conformity to | | IELTS Percentage of agreement |
|-----|---|----------------------------|----|--|------------------------|----|-------------------------------------|
| | | DA | SA | | DA | SA | |
| 1 | The assessment of current independent ability or performance | | ✓ | 100% | | ✓ | 100% |
| 2 | The focus on the whole testee by taking non-intellective factors into account | ✓ | | 60% | ✓ | | 60% |
| 3 | Comparison with peers | | ✓ | 50% | | ✓ | 60% |
| 4 | Emphasis on end- products | | ✓ | 100% | | ✓ | 100% |
| 5 | Formal and standardized testing context | | ✓ | 100% | | ✓ | 100% |
| 6 | Norm-based evaluation format | | ✓ | 100% | | ✓ | 80% |
| 7 | The objective evaluation of receptive skills | | ✓ | 90% | | ✓ | 90% |
| 8 | The neutrality of the examiner | | ✓ | 100% | | ✓ | 80% |
| 9 | Task designs based on psychometric properties | | ✓ | 100% | | ✓ | 100% |
| 10 | Mere assessment through fixed questions | ✓ | | 100% | ✓ | | 100% |
| 11 | No feedback on strengths and weaknesses | | ✓ | 100% | | ✓ | 100% |

DISCUSSION

Discussion of the First and Second Research Questions

The first two research questions were whether the EFL teachers perceived that the productive and receptive sections of the TOEFL iBT and the IELTS observe the DA criteria. The results depicted a few DA traits found in these exams.

One of the DA characteristics attributed to the TOEFL iBT productive sections by the participants was their dominantly subjective scoring procedures. However, while subjective evaluation was overrated in the quantitative phase, the qualitative results showed slightly subjective assessment and the dominance of objective and static assessment in the TOEFL iBT sections. This DA characteristic was also attributed to the IELTS productive sections owing to raters' assessment benchmarks despite the IELTS written evaluation rubrics. In this respect, unlike the quantitative results, the

interview results emphasized the dominantly objective and static interpretation of results.

Another DA characteristic is the prediction of test-takers' ZPD through intervention. Whereas the quantitative results emphasized this dynamic trait in the IELTS productive sections, the qualitative findings disconfirmed it since testees only receive few *graduated hints*, if any, in part 3 of the IELTS interview through dyadic interaction. The qualitative and, somehow, quantitative results accentuated the violation of this DA tenet in the IELTS writing section since visual representations presented in task 1 of the Academic IELTS are not based on testees' needs during the assessment in a *graduated* form. Thus, their learnability cannot be evaluated. As in the IELTS, the quantitative results emphasized the observation of this dynamic feature in the TOEFL iBT productive sections whilst the qualitative ones did not confirm it.

Further, the qualitative and, to some degree, quantitative results stressed the existence of a few DA traces in the third part of the IELTS because of the two-way examiner-testee interaction paving the way for a slightly individualized and flexible examining process via offering some during-the-assessment *graduated* feedbacks based on testees' needs. However, such a feature was not attributed to the TOEFL iBT exam.

Discussion of the Third and Fourth Research Questions

The third and fourth research questions were if the EFL teachers perceived that the productive and receptive sections of the TOEFL iBT and the IELTS observe the SA criteria. In this regard, the study results depicted the observation of most SA features and only a few violations from SA tenets in these exams.

Two of the distinguished SA traits are the comparison of test-takers' skills with those of others and the projection of their future success without any *graduated* feedback. According to the results, these SA features are featured in both sections of the TOEFL iBT since testees either receive no prompts as in the receptive sections and independent parts of its productive sections or get some *non-graduated hints* during the assessment of its integrated sections proving that their learnability cannot be assessed. The same applies to the IELTS writing and receptive sections where testees do not receive any *graduated feedback* during the assessment because of not having face-to-face interaction with evaluators. Although in the third part of the IELTS interview, the interactive context allows for some DA feedback, the SA features in question are dominant in the whole IELTS interview. It is worth mentioning that from the participants' viewpoints in the study qualitative phase, the TOEFL iBT only reports testees' raw scores in each skill and the total of all raw scores but not their percentile rank as in its paper-based counterpart which raises doubt on the existence of these SA traits in the TOEFL iBT receptive sections.

Another significant SA trait observed in the receptive sections of both exams is the objective assessment and interpretation of testees' skills based on pre-determined criteria to boost the result reliability too.

Static exams enjoy a standardized testing context for all testees regardless of their individualized needs during the assessment. According to the results, this SA feature exists in all TOEFL iBT sections, where the same computerized setting is used for the assessment of all candidates, as well as in the IELTS Listening, Reading and Writing Sections.

Discussion of the Fifth Research Question

The last research question of this study was whether the teachers' perceptions show any statistical distinctions between the DA and SA criteria observation in the productive and receptive sections of these two exams. According to the quantitative results, neither the productive sections of the IELTS and TOEFL iBT nor their receptive sections had much compliance with DA principles. Moreover, based on the qualitative results both sections of the exams had more conformity to SA tenets and merely a few traces of DA characteristics are observed in them. Nevertheless, based on EFL teachers' views, the IELTS productive sections are a trifle more in alignment with the DA tenets than those of the TOEFL iBT. One line of explanation is that despite the pre-determined rubrics available for the evaluation of the productive sections of both exams, they are slightly open to the subjective assessment based on examiners' criteria perceived as a DA feature. However, this subjectivity can be a little more dominant in part 3 of the IELTS interview owing to the face-to-face interaction setting the stage for a slightly individualized examining process that is responsive to test-takers' needs while doing the task and focusing on the whole test-takers rather than their isolated competencies. By contrast, the entirely computerized nature of assessment in the TOEFL iBT prevents such dyadic interaction and dynamic assessment.

Another justification for the slight conformity in question might be that the examiner-examinee relationship is interactive in dynamic assessment to examine testees' potential success through *graduated* feedback. According to the results, the examiner-examinee relationship is only real-life and interactive in the IELTS interview. This feature is entirely missing in all sections of the TOEFL iBT owing to its computerized and non-dyadic administration.

A justification for the obtained results of the study is that in spite of the unavoidable advantages of dynamic assessment over static assessment as well as the criticisms leveled at static assessment, testing organizations in charge of the IELTS and TOEFL iBT have not made any modifications in the format of these tests based on DA tenets. The modifications are, however, hard to make due to some logical reasons including critical restraints such as psychometric issues, high costs and the long time needed to carry out dynamic-based exams.

CONCLUSION

The results of the study demonstrated that the IELTS and the TOEFL iBT have few DA features and mostly comply with SA tenets. However, these exams can be modified based on DA principles to offer a better image of testees' current language skills as well as their future learnability. The qualitative results imply that the integrated part of the TOEFL iBT speaking section can be modified according to the DA principles. This can be done by emphasizing testees' learnability through giving them grammatical, lexical or general hints during the oral assessment by real evaluators each time they make a mistake and also by considering non-intellective factors affecting their

performance. If done, dynamic-based integrated tasks and static-based independent tasks in the TOEFL iBT speaking section can evaluate testees' oral skill more optimally and determine their ZOA as well as ZPD more efficiently. Moreover, the IELTS speaking section can also be evaluated dynamically, which can offer a clearer appraisal of testees' oral and communicative skills.

This study puts forward some implicit suggestions for the organizations in charge of the design, administration, and supervision of the IELTS and the TOEFL iBT exams to modify their test contents by designing some dynamic-based tasks in some sections of these high-stakes exams. Simply put, IELTS and the TOEFL iBT test designers and decision-makers may find the findings of this study useful and valuable. Likewise, the results of this study may propel psychometricians, policy-makers and developers of standardized exams in general and the IELTS and the TOEFL iBT exam designers in particular to revise and refine these high-stakes test designs and administration processes based on the integrated features of dynamic and static assessments in the near future to mirror a more accurate image of the current language proficiency and the future developmental potential of test-takers. Furthermore, the results of this study can help TOEFL iBT and the IELTS test designers look at these exams in a different light and alter them accordingly.

Hopefully, the study results will raise researchers and standardized exam test-makers' cognizance of the potentials of dynamic assessment to be utilized in high-stakes exams and motivate them to conduct more research in this area.

References

- Ajideh, P., & Nourdad, N. (2012). The effect of dynamic assessment on EFL reading comprehension in different proficiency levels. *Language Testing in Asia*, 4(3), 101.
- Bingham, W. V. D., & Moore, B. V. (1959). How to interview. *New York: Harper & Row*.
- Ebadi, S., & Rahimi, M. (2019). Mediating EFL learners' academic writing skills in online dynamic assessment using Google Docs. *Computer Assisted Language Learning*, DOI.
- Grigorenko, E. L., & Sternberg, R. J. (1998). Dynamic testing. *Psychological Bulletin*, 1(124), 75-111.
- Haywood, H. C., & Lidz, C. (2006). Dynamic assessment in practice. Clinical and educational applications. . *New York: Cambridge University Press*.
- Khoshsima, H., & Izadi, M. (2014). Dynamic vs. standard assessment to evaluate EFL learners' listening comprehension. *Iranian Journal of Applied Language Studies*, 2(6), 1-26.
- Kozulin, A., & Garb, E. (2002). Dynamic assessment of EFL text comprehension of at-risk students. *School Psychology International*, 1(23), 112-127.
- Merrell, K. W., Erwin, R. A., & Peacock, G. G. (2012). *School Psychology for the 21st century: Foundations and practices*. . *New York, NY: Guilford Press*.
- Morrison, T. (2001). *Actionable learning: A handbook for capacity building through case-based learning*. Tokyo. *Asian Development Bank Institute*.
- Schwab, D. P. (1980). Construct validity of organizational behavior. In B.M. Staw & L.L. Cummings (Eds.) *Research in organizational behavior*, 1(2), 3-43.
- Tzuriel, D. (2001). *Dynamic assessment of young children*. New York. *Kluwer Academic, Plenum Publishers*.
- Tzuriel, D., & Klein, P. S. (1987). Assessing the young child: Children's analogical thinking modifiability. In C.S. Lidz (Ed.), *Dynamic assessment: An interactional approach to evaluating learning potential*, 268-287.
- Vygotsky, L. S. (1998). The problem of age. In R. W. Rieber (Eds.). . *The collected works of L. S. Vygotsky: Child psychology*, 187-205.

Biodata

Ms Arezoo Daneshvar is a Ph.D. candidate at the Islamic Azad University, Shiraz branch, Iran. Her areas of interests include teaching English to advanced EFL learners, doing research to discover unknown issues in her field of study, boosting her knowledge of German and other foreign languages to do comparative and contrastive analysis of languages.

Email: a.daneshvar20@gmail.com

Dr Mohammad Sadegh Bagheri is a PhD of TEFL, a member of TESOL Arabia, ACLA and Shiraz Islamic Azad University Research Committee, as well as a full-time faculty member of EFL Department, Islamic Azad University, Shiraz Branch, Iran. He has published more than sixty articles and forty books and presented papers in several national and international conferences. His most famous book is Crack IELTS series and his research areas are international proficiency exams, research, teaching methods and assessment.

Email: Bagheries@gmail.com

Dr Firooz Sadighi, is a PhD of English language and linguistics at Shiraz University. He got retired in 2009. Currently, he is actively engaged in teaching different courses at the Islamic Azad University, Shiraz Branch, Iran. He has published numerous articles nationally and internationally, and a number of books independently and collaboratively with colleagues and students. His research areas of interest include 'foreign/second language learning and teaching' and 'linguistics'. He has also presented several papers at local and international conferences.

Email: firoozsadighi@yahoo.com

Dr Lotfollah Yarmohammadi, is a PhD of linguistics. He is now retired but he is currently teaching at the English Department at Islamic Azad University, Shiraz, Iran. He has taught native and non-native speakers of Persian in Iran and overseas. His research areas of interest include 'pragmatics' and 'linguistics'. He has written and published several papers and books on linguistics in national and international journals.

Email: Yarmohmal@yahoo.com

Dr Mortaza Yamini, is a PhD of TEFL, and a faculty member of EFL Department, Shiraz University, Iran. He is now an academic member of Zand Institute of Higher Education. His research interests include language teaching methodology, testing and phonology. He has extensively published in local and international journals and has authored several books.

Email: mortazayamini@gmail.com or yamini@zand.ac.ir