Research Paper

# TOEFL iBT Integrated and Independent Writing Tasks: Iranian Test Takers' Performance in Focus

**Farah Shooraki[1], Hossein Barati[2*], Ahmad Moinzadeh[3]**

[1]Ph.D. Candidate, Department of English, Maybod Branch, Islamic Azad University, Maybod, Iran

[2*]Associate Professor, Department of English Language and Literature, University of Isfahan, Isfahan, Iran

[3]Associate Professor, Department of English Language and Literature, University of Isfahan, Isfahan, Iran

## Abstract

This study aimed to compare the performance of international and Iranian test takers on TOEFL iBT integrated and independent writing tasks. The international test takers' data were provided by TOEFL organizer: Educational Testing Service (ETS). This included a total of number of 4916 samples for both integrated and independent tasks. A paired sample t-test showed the international test takers significantly outperformed on the independent task. Further, a sample of 100 integrated and independent writing tasks produced by international test takers was randomly selected through systematic sampling, and compared with 96 counterparts written by Iranian TOEFL iBT test takers. Although the difference between Iranian and International test takers was not significant, the paired sample t-test on Iranian samples per se showed that they too outperformed on the independent writing tasks.

**Keywords:** Coh-Metrix, independent writing task, integrated writing task, TOEFL-iBT, writing assessment

## INTRODUCTION

In large scale testing situations such as TOEFL, independent writing has been widely used as a measure of second language (L2) academic writing ability. It is generally agreed that compared with indirect writing assessment (e.g., multiple-choice questions), independent writing tasks provide a more valid representation of underlying writing ability because they afford the assessment of writing performance beyond morphological and syntactic manipulation (Camp, 1993). Unlike indirect writing assessments, independent writing tasks prompt test takers to produce an extended written argument built exclusively on their prior knowledge and/or experience.

ETS (2022) asserts that the writing section of TOEFL iBT which takes a total of 50

**\*Corresponding Author's Email:**
*barati@fgn.ui.ac.ir*

minutes, measures learners` ability to communicate clearly in writing and compose well-organized essays using correct grammar, spelling, vocabulary, and sentence structure in an academic setting. It consists of two types of tasks: The Reading-Listening-Writing Integrated Task and Independent Writing Task.

The first is an integrated task that requires test takers read a passage on an academic topic and then listen to a speaker discuss it. They, then, write a summary about the important points in the listening passage and explain how these relate to the key points of the reading passage. However, in the independent section, the test takers are given one topic, and they have 30 minutes to write an essay of about 4-5 paragraphs, or 300-350 words. They must write an essay that states their opinion or choice, and then explain it, rather than simply listing personal preferences or choices (ETS,

2022). It should be mentioned that previously, the TOEFL writing component contained only an independent task. However, this task was criticized because it did not closely resemble the genres used in real academic settings.

The conceptualization of writing in TOEFL iBT is based on the TOEFL 2000 and Cumming, Kantor, Powers, Santos, and Taylor (2000) frameworks, which focus on testing writing in academic settings (Riazi, 2016). "As conceptualized by Cumming et al. (2000), writing in academic settings typically involves the production of written text forms to meet certain expectations to complete course-related assignments and to display the knowledge they are acquiring" (p: 16). Based on this conceptualization, the assumption then is that, if test takers were able to perform well on TOEFL iBT writing test tasks, they would be able to use the core writing abilities to perform discipline-specific assignments. The underlying goal of including both writing tasks is to enhance the authenticity and validity of English as a Second Language (ESL) writing tests (Cumming et al., 2005, 2006; Stricker & Attali, 2010).

However, concerns have been raised about writing assessments and learning that solely contain independent writing tasks because they risk decontextualizing the writing activity and obtaining an unrepresentative snapshot of the writer's abilities (Barkaoui, 2015; Oxford, 2006; Plakans & Gebril, 2013). As a consequence, independent writing assessments may under-represent writing proficiency (Cumming, 2013; Kim, 2017; Riazi, 2016). To address these concerns, integrated writing tasks have been proposed as a promising alternative for standardized writing tests such as TOEFL iBT (Cumming et al., 2005; Plakans, Gebril, & Bilki, 2019; Plakans & Gebril, 2013).

In terms of testing validity, the combined use of integrated and independent writing tasks can diversify and improve overall measures of writing ability because no single task can be solely reliable to predict the writing ability of a test-taker (Cumming, 2013; Plakans et al., 2019). The use of integrated writing tasks prompts test-takers to respond to source text(s), testing their ability to identify and extract relevant information in the source text(s), and organize

and synthesize the information in the response they construct (Cumming et al., 2000). Independent writing tasks, on the other hand, prompt test-takers to produce an extended written argument built exclusively on their prior knowledge and/or experience. Because of the source materials and the academic nature of integrated writing tasks, integrated essays are expected to be different from independent essays and contain more sophisticated language forms (Cumming et al., 2005, 2006; Plakans & Gebril, 2012).

Despite the many advantages of integrated writing tasks, there have been relatively few studies on these tasks in the literature of L2 writing assessment, especially when compared with the abundance of research on independent writing tasks. Given that other modalities of communication (such as reading and/or listening) are involved in integrated writing, questions have been raised about what such tasks really tap into and whether use of such tasks increases risks of confusing assessment of comprehension with assessment of writing ability (Plakans et al., 2019).

The inclusion of integrated writing tasks in TOEFL-iBT aimed to overcome the shortcomings identified in the validation studies of independent writing tasks for writing assessment (Cho, Rijmen & Novak 2013; Cumming et al. 2006; Plakans, Gebril 2021). Although there is an extensive amount of research available on independent writing tasks (Grant & Ginther 2000; Plakans et al., 2019), examinations of L2 writing assessment reveal that only a few studies have focused on integrated writing tasks (e.g., Guo et al. 2013; Gebril 2021; Plakans et al., 2019).

Cumming et al. (2000) claimed that writing in TOEFL-iBT typically involves the production of written text forms through which students need to complete course related assignments and display the knowledge they have acquired so far. Cumming et al. then conceptualized that "text characteristics have been derived from specific sets of texts written by particular populations in particular contexts, so their potential to generalize to other texts, populations, or contexts is probably quite limited" (2000:16). This conceptualization indicates

that none of these text characteristics can be considered thoroughly validated or uniquely integral to the full range of ESL writing. Overall, based on Cumming et al.'s (2000) conceptualization and empirical studies (e.g. Biber & Gray 2013; Cumming et al. 2006; Gebril & Plakan 2009, 2013; Guo et al. 2013; Riazi 2016), the assumption is then that the texts produced by EFL test-takers when completing integrated and independent iBT TOEFL tasks are different in terms of linguistic and discoursal features.

Barkaoui (2015) examined the writing activities that test takers engage in when responding to the writing tasks in the *TOEFL iBT®* test and to examine the effects of task type and test-taker English language proficiency (ELP) and keyboarding skills on the frequency and distribution of these activities. He found out that the participants engaged in various construct-relevant activities, such as interacting with the writing task and resources, planning, generating, evaluating, and revising. Additionally, participants' writing activities varied most across writing tasks and, to a lesser extent, across English proficiency groups.

Gholami and Alinasab (2017) explored the effects of source-based writing practice on EFL learners' composing abilities and investigated the probable differences between those tasks and independent writing ones in improving Iranian EFL learners' essay writing abilities. Based on the findings, the participants with hybrid writing practice outperformed their counterparts in integrated essay tests. Their superior performance was not observed in the case of traditional independent writing tasks.

Aminzadeh and Booheh (2015) attempted to examine the impacts of reading-to-write and writing-only tasks on the writing ability of Iranian EFL learners. Findings revealed a statistically significant difference between the writing ability of the participants of the two groups. In fact, integrated writing task were found to be more effective than independent tasks. The data collected via the reading-to-write questionnaire was examined with the results indicating a preference for reading-to-write tasks.

Soleimani and Mahdavipour (2014) also studied the effect of variations in integrated writing tasks and proficiency level on features of the texts written by Iranian EFL learners. Results of ANOVAs showed that firstly, variations in integrated writing tasks together with level of proficiency had a significant effect on the discoursal features, secondly, the integrated tasks generated features sharing the same features, and thirdly, a certain level of proficiency could be distinguished trough some of these features. Moreover, the findings showed that direct source use without quotation (i.e., plagiarism) is found higher in reading-to-write task than listening-to-write task especially among the low proficiency writers.

Therefore, this literature review calls for more studies on the performance of learners on TOEFL-iBT writing tasks. Consequently, the following research questions were addressed:

*RQ 1: Do test takers perform significantly differently on the independent and integrated writing tasks of iBT TOEFL?*

*RQ 2: Do Iranian and international test takers perform significantly differently on the independent and integrated writing tasks of iBT TOEFL?*

## METHODS
### Design
In line with its objectives, this study benefited from an Ex post facto design. It was Ex post facto research because no manipulation or control of the independent variable was done in this study (Ary, et al., 2019).

### Materials
The data for the first research question came from 2458 international test takers by two independent and integrated writing tasks of TOEFL iBT (4916 writing samples). Whether graduate or undergraduate, the international test takers were adult, male and female, from different first language backgrounds and different disciplines of science and engineering, arts and humanities, business, economics, etc. They came from 130 different countries. A similar range was found in the L1s spoken by participants. Participants selected more than 67 different L1s. Among the test takers, 49% were females, 50% were males, and the remaining 1%

of the test takers did not supply their gender. The test takers' scores and information were provided by ETS, TOEFL iBT Public Use Dataset - Writing section dataset. The data came from different administrations since 2006. This was a very rich source of data to analyze and also to find particular issues that were later observed in the analysis enriching the knowledge about the test takers.

**Table 1**
*Percentage of Test-Takers*

|         |        | Frequency | Percent | Valid Percent | Cumulative percent |
|---------|--------|-----------|---------|---------------|--------------------|
| Valid   | Female | 1043      | 42.4    | 49.3          | 49.3               |
|         | Male   | 1071      | 43.6    | 50.7          | 100                |
| Missing | Total  | 2114      | 86.0    | 100           |                    |
|         | System | 34414     |         |               |                    |
| Total   |        | **2458**  | **100** |               |                    |

The mean score of respondents' age was 23.53 (SD = 6.49) with Minimum 10 and Maximum 57. The mean score of respondents' integrated writing task was 2.98 (SD = 1.22) with Minimum 1.00 and Maximum 5.00. The mean score of respondents' independent writing task was 3.36 (SD = .85) with Minimum 1.00 and Maximum 5.00.

The data for the second research question was also provided by ETS. The Iranian test takers were 48 adults, male and female, and from different educational majors. Farsi was their first Language. The test takers' mean age was 29.93 years (SD = 6.429). The youngest one was 19 years old and the oldest was 47 years old. Among the test takers, 36.4% were females, 63.6% were males (see figure 3.2). Also, a sample of 100 international test takers' essays including both integrated and independent writing tasks produced by 50 test takers worldwide was randomly collected through systematic sampling.

**Instruments**
In this study, both the writing tasks (i.e., integrated and independent tasks) and the writing rubrics of ETS were used as main tools. That is, the test takers' samples of written discourses called *writing tasks* provided the researcher with the scope of an objective assessment to trace test takers' actual writing development. The writing test tasks used in the study were provided by ETS and consisted of authentic test materials. For scoring, the writing tasks were scored based on the Writing Scoring Guides (Rubrics) by a combination of AI scoring and certified human raters. Raw scores are converted to a scaled section score of 0–30.

**Integrated Task**
In the integrated task, test-takers were supposed to read a passage on an academic topic and then listen to a speaker discuss it. The lecture usually takes a position that is somehow different from the position presented in the reading passage. The test-takers then, in connected English prose, wrote a summary about the important points in the listening passage and explained how those related to the key points of the reading passage. That is, they were given 20 minutes to write an essay of about 150-225 words, explaining how the listening passage either supported the reading passage or contradicted the reading. The integrated essay was evaluated holistically on its organization, appropriate and precise use of grammar and vocabulary, and completeness and accuracy of the content covered in the source materials (ETS, 2022).

**Independent Task**
In the independent task, however, test takers received no oral or written stimulus materials; instead, they were given one topic; they had 30 minutes to write an essay of about 4-5 paragraphs, or 300-350 words. For this task, they responded to a relatively general question that allowed them to tap their own knowledge and experience. They were supposed to write an essay that stated, explained, and supported their opinion on an issue, and develop support for their opinions rather than simply listing

personal preferences or choices. The independent essay was graded holistically based on the development, organization, and appropriate and precise use of grammar and vocabulary of their writing (ETS, 2022).

## Writing Rubrics

ETS (https://www.ets.org/s/toefl/pdf/toefl_writing_rubrics.pdf, 2022) asserts that the both integrated and independent rubrics describe five levels of writing performance. That is, both tasks are rated from 0 to 5. The sum is converted to a scaled score of 0 to 30.

| Writing | 0–30 | Good | 24–30 |
| | | Fair | 17–23 |
| | | Limited | 1–16 |

For the integrated task, the scoring rubric principally focuses on accurate and coherent presentation of the extracted information in the essays in addition to grammatical accuracy. An integrated essay with a score of 5 should successfully select important ideas from source text(s), present them coherently and accurately, and contain only minor and occasional language errors. An integrated essay with a score of 1, however, provides either little or no relevant content from the source text(s) or is difficult to understand due to the severity of language mistakes.

In the independent scoring rubric, linguistic sophistication at the lexical and syntactic levels is emphasized in addition to the logic and coherence of the arguments along with grammatical accuracy. An independent essay with a score of 5 should be a well-organized and developed response to the given topic, displaying linguistic sophistication and containing only minor language mistakes. An essay with a score of 1, on the other hand, has serious problems in organization, idea development, or language use.

ETS uses both human raters and automated scoring methods to offer a complete and accurate picture of a test taker's ability.

Human rating — multiple, rigorously trained raters score tests anonymously. ETS raters are continually monitored to ensure fairness and the highest quality.

eRater®- automated scoring technology is used with human ratings to score the independent and integrated writing tasks.

Using both human judgment for content and meaning with automated scoring for linguistic features ensures consistent, quality scores (ETS, 2022).

## Data Collection Procedure

To analyze the first research question, the TOEFL-iBT test takers integrated and independent writing tasks and scores obtained from ETS were analyzed using SPSS (Edition 22) software. The data included 4916 writing essays taken from 2458 international test takers of TOEFL iBT. The essays had been marked by ETS raters including human raters and e-rater automated scoring technology. They used the TOEFL writing scale, which ranged from 0 to 5. The sum was converted to a scaled score of 0 to 30. A final score was available for each of the integrated and the independent essays. A paired sample t-test was administered to assess whether the international test takers performed significantly differently on the independent and integrated writing tasks on iBT TOEFL.

To address the second research question, the data was provided by ETS. That is, a sample of 100 test takers' essays including both integrated and independent writing essays produced by 50 test takers worldwide was randomly collected through systematic sampling and compared with the sample of 96 essays including both tasks produced by 48 Iranian test takers of TOEFL iBT. Both the integrated and the independent writing tasks were graded by ETS raters including human raters and e-rater automated scoring technology. They used the TOEFL writing scale, which ranged from 0 to 5. The sum was converted to a scaled score of 0 to 30. A final score was available for each of the integrated and the independent essays. A paired sample t-test, therefore, was run to compare the performance of Iranian and international test takers of TOEFL iBT on both the integrated and the independent writing tasks.

## RESULTS

This study aimed to analyze the performance of international and Iranian test takers of TOEFL iBT in response to independent and integrated writing tasks.

## Analysis of the Performance of International

## Test-Takers of TOEFL iBT on the Integrated and Independent Writing Tasks

Table 2 shows the t-test results for integrated and independent writing tasks. Since the size of T is less than 0.05, there is a significant difference between the integrated and independent writing tasks (P< 0.001).

**Table 2**

*Paired Samples Test Results for Integrated and Independent Writing Tasks*

| Group | Mean | N | Std. Deviation | T | P |
|---|---|---|---|---|---|
| Integrated | 2.98 | 2458 | 1.22 | -20.234 | .001 |
| Independent | 3.36 | 2458 | .85 | | |

Table 3 presenting the descriptive statistics for test-takers' age as well as integrated and independent writing scores were examined to avoid any hunches.

**Table 3**

*Descriptive Statistics of Test-Takers: Age and Integrated and Independent Writing Scores*

| | Minimum | Maximum | Mean | Std. Deviation |
|---|---|---|---|---|
| Age | 10 | 57 | 23.53 | 6.49 |
| Integrated | 1.00 | 5.00 | 2.98 | 1.22 |
| Independent | 1.00 | 5.00 | 3.36 | .85 |

In order to verify the relationship among age, integrated and independent writing tasks, Pearson coefficient was used. Table 4 shows that the variables were not meaningfully related to age.

**Table 4**

*Correlation Coefficients among Age, Integrated and Independent Writing Task*

| | | Integrated | Independent |
|---|---|---|---|
| **Age** | Pearson Correlation | .014 | .015 |
| | Sig. (2-tailed) | .503 | .457 |

Table 5 indicates that there is no significant difference between integrated writing task and gender.

**Table 5**

*Independent t-Test Results for the Integrated Writing Task and Gender*

| Gender | N | Mean | Std. Deviation | T | P |
|---|---|---|---|---|---|
| Male | 1071 | 2.97 | 1.24 | .62 | .53 |
| Female | 1043 | 3.0 | 1.20 | | |

The results shown in Table 6 indicate that there is no significant difference between independent writing score and gender (p>.05).

**Table 6**

*Independent t-Test Results for the Independent Writing Task and Gender*

| Gender | N | Mean | Std. Deviation | t | P |
|---|---|---|---|---|---|
| Male | 1071 | 3.35 | .88 | 1.514 | .130 |
| Female | 1043 | 3.40 | .83 | | |

Table 7 illustrates the correlation coefficients between year and mean score of writings. The results show there is a positive and significant relationship between year and mean score of writing (correlation coefficient = 0.894).

**Table 7**
*Correlation Coefficients between Year and Mean Score of Writings*

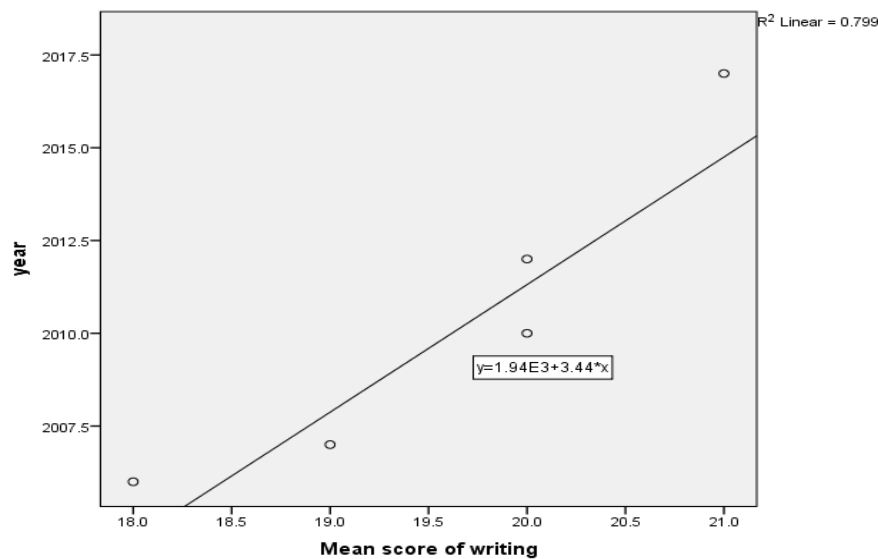|  |  | year | Mean score of writings |
|---|---|---|---|
| year | Pearson Correlation | 1 | .894* |
|  | Sig. (2-tailed) |  | .016 |
| Mean score of writing | Pearson Correlation | .894* | 1 |
|  | Sig. (2-tailed) | .016 |  |



**Figure 1**
*Chart of scatter plot correlation between year and Mean score of writings*

In sum, tables so far showed that the international test takers of TOEFL iBT performed significantly differently on independent and integrated writing tasks. Interestingly enough, they outperformed on the independent writing tasks than the integrated tasks. Also results indicated that there was no significant difference between international male and female performance on integrated and independent writing tasks (r= 0.89). Therefore, gender and age were not considered critical and influential factors in the written performance of TOEFL iBT test takers. Finally, findings showed that there was a positive and significant relationship between the years of administrating TOEFL iBT and the mean score of international test takers. In other words, the performance of test takers on independent and integrated writing tasks has increased over the years, i.e., from 2007 to 2017.

**Comparison of the Performance of Iranian and International Test-Takers of TOEFL iBT on the Integrated and Independent Writing Tasks**
Before running the independent samples t-test, the common assumptions for the analysis of a t-test regarding the scale of measurement, normality of data distribution, adequacy of sample size and equality of variance in standard deviation were checked. For the scale of measurement, each test-taker's performance on each writing task was assessed based on TOEFL-iBT integrated and independent writing. Also, normality of data distribution was checked in SPSS using the Shapiro-Wilk normality test. Table 8shows the data are normally distributed.

**Table 8**

***Shapiro-Wilk Normality Test for the Iranian Integrated and Independent Writing Tasks***

| Group | Shapiro-Wilk | | |
|---|---|---|---|
| | Statistic | df | Sig. |
| Integrated | .957 | 24 | .236 |
| Independent | .929 | 24 | .086 |

Table 9 shows that there is no significant difference between the integrated and independent writing tasks (P< 0.001). Although the difference is not significant, they outperformed on independent writing tasks than integrated tasks.

**Table 9**

***Paired Samples Test Results for the Iranian Integrated and Independent Writing Tasks***

| Group | Mean | Std. Deviation | t | P |
|---|---|---|---|---|
| Integrated | 3.03 | .83 | -1.05 | .29 |
| Independent | 3.16 | .77 | | |

Like the first research question, age, gender and year variables, though not the determining factors in this section, were examined to avoid any hunches. Table 10 presents the descriptive statistics for test-takers' age, as well as integrated and independent writing scores.

**Table 10**

***Descriptive Statistics of the Iranian Test-Takers: Age and Integrated and Independent Writing Scores***

| | Minimum | Maximum | Mean | Std. Deviation |
|---|---|---|---|---|
| Age | 19 | 47 | 29.93 | 6.42 |
| Integrated | 1.0 | 4.5 | 3.03 | .83 |
| Independent | 1.5 | 4.5 | 3.16 | .77 |

Table 11 shows in order to verify the relationship among age, integrated and independent writing tasks, Pearson coefficient was used. The variables were not meaningfully related to age.

**Table 11**

***Correlation Coefficients between Age, Integrated and Independent Writing Tasks of Iranian***

| | | Integrated | Independent |
|---|---|---|---|
| Age | Pearson Correlation | -.169 | -.098 |
| | Sig. (2-tailed) | .390 | .622 |

The results of Table 12 show that there is no significant difference between integrated writing task and gender (p>.05).

**Table 12**

***Independent t-Test Results for the Iranian Integrated Writing Task and Gender***

| Gender | Mean | Std. Deviation | t | P |
|---|---|---|---|---|
| Male | 3.16 | .85 | 2.01 | .05 |
| Female | 2.65 | .70 | | |

The results shown in Table 13 indicate that there was not a significant difference between independent writing score and gender (p>.05).

**Table 13**

*Independent t-Test Results for the Iranian Independent Writing Task and Gender*

| Gender | Mean | Std. Deviation | t | P |
|--------|------|----------------|-----|-----|
| Male | 3.21 | .78 | .47 | .63 |
| Female | 3.09 | .84 | | |

**Table 14**

*Independent t-test results for Iranian and International Independent Writing Task*

| Test Takers | Mean | Std. Deviation | t | P |
|-------------|------|----------------|-------|-----|
| Iranian | 3.16 | .77 | -1.02 | .31 |
| International | 3.34 | .95 | | |

**Table 15**

*Independent t-Test Results for Iranian and International Integrated Writing Task*

| Test Takers | Mean | Std. Deviation | t | P |
|-------------|------|----------------|------|------|
| Iranian | 3.032 | .8302 | .418 | .677 |
| International | 2.940 | 1.2961 | | |

**Table 16**

*Independent t-Test Results for Total Score of Iranian and International Integrated Writing Task*

| Test Takers | Mean | Std. Deviation | t | P |
|-------------|------|----------------|-------|-----|
| Iranian | 6.191 | 1.3735 | -.247 | .80 |
| International | 6.280 | 2.1048 | | |

In sum, the second research question aimed to compare the performance of Iranian and international test takers of TOEFL iBT writing section on both independent and integrated writing tasks. The findings revealed that there was no significant difference between the integrated and independent writing tasks (P< 0.80). In other words, Iranian and international test takers of TOEFL iBT did not perform significantly differently on independent and integrated writing tasks. Also results showed that there was no significant difference between Iranian and international male and female performance on integrated and independent writing tasks (p>.05). Therefore, gender and age were not considered critical and influential factors in the written performance of TOEFL iBT test takers as well (see Tables 15 and 16).

## DISCUSSION

This study analyzed the actual performance of international and Iranian test-takers on integrated and independent writing tasks as part of a broader project regarding the validation of TOEFL-iBT. The study, similar to Cumming et al. (2006), suggests that the new TOEFL writing section has been successful in eliciting more than one type of discourse from the writers. The results also provide empirical evidence for the TOEFL validity argument for the combined use of integrated and independent writing tasks) in that different language skills are evaluated by the two tasks.

The first aim of the study was to investigate the performance of international test takers of TOEFL iBT writing section on both independent and integrated writing tasks. The findings showed that there was significant difference between the integrated and independent writing tasks. So, the null hypothesis that there is no significant difference between test takers in performing on the independent and integrated writing tasks of iBT TOEFL was rejected. Interestingly enough, they outperformed on independent writing tasks than integrated tasks.

The results corroborate findings reported in previous studies done by Barkaoui (2015), and Cumming et al. (2006) who claimed that integrated tasks are "more complex and more

demanding than traditional stand-alone or independent tasks, in which test-takers draw on their own knowledge or ideas to respond to a question or prompt".

However, the results of the study are not in line with the studies by Gebril et al. (2009), who also found a high correlation between the two sets of scores when comparing the performance of a group of EFL students on a thematically-related integrated task and an independent writing task.

The second aim of the study was to examine the performance of Iranian test takers of TOEFL iBT writing section on both independent and integrated writing tasks. The results showed that Iranian test takers of TOEFL iBT did not perform significantly differently on independent and integrated writing tasks. Although the difference was not significant, the finding per se indicated that the test takers too outperformed on independent writing tasks than integrated tasks. Also results showed that there was no significant difference between Iranian male and female performance on integrated and independent writing tasks (p>.05). Therefore, gender and age were not considered critical and influential factors in the written performance of Iranian TOEFL iBT test takers. So the null hypothesis that there is no significant difference between test takers in performing on the independent and integrated writing tasks of iBT TOEFL is accepted.

This study is not consistent with the studies by Aminzadeh and Booheh (2015) who found that there was a statistically significant difference between the writing ability of the participants of the two groups. They examined the impacts of reading-to-write and writing-only tasks on the writing ability of Iranian EFL learners. In fact, integrated tasks were found to be more effective than writing-only tasks. The data collected via the reading-to-write questionnaire was examined with the results indicating a preference for reading-to-write tasks.

Similarly, this study is also not in harmony with the study by Gholami and Alinasab (2017) who found that test takers with hybrid writing practice outperformed integrated essay tests. Esmaeili (2002) also compared thematically-related integrated writing scores with independent writing scores and found that the

ESL participants achieved significantly higher scores in the integrated than in the independent writing task (F = 134.28, p = .001). The results suggested that when there is a thematic link between different writing activities, the writing scores improved significantly as compared to those in activities without a thematic link.

## CONCLUSION

Integrated writing tasks were included in the new TOEFL to improve the validity, authenticity, and wash back effects of English as a second language (ESL) writing test (Cumming et al. 2006). Given that TOEFL iBT plays a critical role in determining ESL/EFL test takers' admission and placement in college study, understanding of the integrated writing task is fundamental to the design, development, and use of the test. In this way, test takers have the benefit of practicing different language skills in an integrated, natural, communicated way even if one skill, for example, writing is the main focus.

The current study, therefore, shows that the combined use of the two tasks broadens the representation of the underlying academic writing ability and thus provides justification for the addition of the integrated writing task in the writing test. Furthermore, the descriptive information about the specific linguistic performance and writing behaviors associated with the integrated and the independent writing tasks also help to shed light on the construct inherent in each of the tasks. The integrated skill approach exposes English language learners and test takers to authentic language and challenges them to interact naturally in the language. In this way, test takers understand the complexity and richness of the English language for communication purposes. Moreover, this approach shows that English as an international language is not just an object of academic interest or only a key to passing an examination; in fact, English becomes a true means of communication. It can be highly motivating to test takers of all ages and backgrounds. With careful reflection and planning they can use English effectively for communication.

It is assumed this research can confirm that the writing section of TOEFL iBT is greatly developed when test takers inculcate interest in

language learning goals which are mediated in this case through the different tasks. Also, the test takers writing may shed light on their own application of writing strategies evidenced by the different written discourses. Based on test takers performance on the two writing tasks as well as the effectiveness of using rubrics, the attempt is to evaluate whether and how the discourses written for integrated-tasks differ from the discourses written for independent essays. Instructional strategies can also be suggested so that teachers can adopt a more effective approach to enhance students' writing proficiency for TOEFL iBT.

The results also hold a number of implications for L2 writing assessment, and L2 writing instruction and learning. As for L2 writing assessment, the findings show that the holistic scores on writing tasks of EFL test-takers are to be subject to verification and can be anchored empirically via the analysis of discourse qualities like cohesion. In this way, the information can help stakeholders of the test to verify test constructs in terms of particular contexts like EFL and communicative views of language proficiency, and to develop reasonable writing tasks and rating scales. Developing such tests can serve as a basis for the use and interpretation of iBT TOEFL scores. Accordingly, TOEFL test-takers will also benefit from the test.

As for L2 writing instruction, L2 learners and teachers who want to improve on academic writing ability for higher education should be exposed to both types of writing tasks. The integrated tasks provide learners with an opportunity to demonstrate text qualities that are different from those which are demonstrated during independent tasks. Offering lexical items, moreover, within a wide variety of settings, domains, and genres seems beneficial. In order to improve the quality of teaching and assessment, teachers also need useful and easily accessible tools such as Coh-Metrix. It is hoped that teachers make use of reliable computational programs like Coh-Metrix to analyze the students' writing tasks.

In sum, educators, policy makers, and experts realize that minimum and basic skills are no longer sufficient and are calling for a closer match between the skills students learn in school/colleges and the skills they will need upon leaving school or college campus. Schools/universities are now expected to help students develop skills and competencies in real-life or authentic situations, and they are expected to graduate students who can demonstrate these abilities, often measured by alternative assessments rather than traditional pencil-paper tests. Thus, in addition to the assessment of the students writing performance on the basis of scoring rubrics, this research throws some light on the raters/instructors of writing who are to make their decision on the scores given to test takers on their assigned tasks.

However, the findings of this study should be interpreted within the limitations of the study. The sample of Iranian test-takers whose essays were used for the research was relatively small in number (48). Conducting similar studies with a larger quantity of TOEFL-iBT essays seems inevitable. Moreover, due to the restricted sample size of the third research question, any generalizations about the results should be made with caution.

Future studies can compare the writing performance of Iranian test takers with different levels of L2 proficiency and keyboarding skills, and/or compare the writing performance of Iranian test takers between writing in academic settings and writing in a test situation.

**Refrences**

Aminzadeh, R. and Booheh, Z. S. (2015). The Comparative Effect of Reading-to-Write and WritingOnly Tasks on the Improvement of EFL Learners' Writing Ability. *The Journal of Applied Linguistics,* 6 (12).

Amiryousefi, M., & Tavakoli, M. (2014). An empirical study into the effects of Iranian test takers' personal attributes on their TOEFL scores: Reading, writing, and listening in focus. *International Journal of Research Studies in Language Learning, 3(6),* 35-45.

Barkaoui, K. (2015). Test Takers' Writing Activities During the TOEFL iBT® Writing Tasks: A Stimulated Recall Study. *ETS Research Report No.* RR–15-04.

Biber, D., & Gray, B. (2013). Discourse characteristics of writing and speaking task types on the TOEFL-iBT test: A lexico-grammatical analysis. *TOEFL-iBT research report*. Retrieved from http://www.ets.org/Media/Research/pdf/RR-13-04.pdf

Camp, R. (1993). Changing the model for the direct writing assessment. In M. M. Williamson & B. A. Huot (Eds.), Validating holistic scoring for writing assessment: theoretical and empirical foundations (pp. 45-78). Cresskill, NJ: Hampton Press, Inc.

Campbell, C. (1998). *Teaching second-language writing: Interacting with text*. Boston: Heinle & Heinle.

Cho, Y., Rijmen, F., & Novak, J. (2013). Investigating the effects of prompt characteristics on the comparability of TOEFL iBT integrated writing tasks. *Language Testing,* 30(4), 513-534.

Crossley, S. A., Kristopher, K., & Dascalu, M. (2018). The Tool for the Automatic Analysis of Cohesion 2.0: Integrating semantic similarity and text overlap. *Behavior Research Methods,* 51(1), 14–27.

Cumming, A. (1998). Theoretical perspectives on writing. *Annual Review of Applied Linguistics*, 18, 61-78.

Cumming, A. (2013). Assessing integrated writing tasks for academic purposes: Promises and perils. *Language Assessment Quarterly, 10, 1*, 1-8.

Cumming, A., Kantor, R. Baba, K., Erdoosy, U., Eouanzoui, K., & James, M. (2006). Analysis of discourse features and verification of scoring levels for independent and integrated tasks for the new TOEFL (TOEFL Monograph No. MS-30). Princeton, NJ*: ETS.*

Cumming, A., Kantor, R., Baba, K., Erdosy, U., Eouanzoui, K., & James, M. (2005). Difference in written discourse in independent and integrated prototype tasks for next generation TOEFL. *Assessing Writing, 10*, 5–43.

Cumming, A., Kantor, R., Powers, D., Santos, T., & Taylor, C. (2000). TOEFL 2000 writing framework: A working paper (TOEFL Monograph Series No. 18). Princeton, NJ: Educational Testing Service.

Cumming, A., Grant, L., Mulcahy-Ernt, P., & Powers, D. (2004). A teacher verification study of speaking and writing prototype tasks for a new TOEFL. *Language Testing, 21*(2), 159–197.

Esmaeili, H. (2002). Integrated reading and writing tasks and ESL students' reading and writing performance in an English languagetest.*The Canadian Modern Language Review*, 58, 599–620.

ETS. (2022). TOEFL iBT® Test Content. Retrieved from https://www.ets.org/toefl/ibt/about/content/.

Gebril, A. (2009). Score generalizability of academic writing tasks: Does one test method fit it all? *Language Testing,* 26(4), 507-531.

Gebril, A., & Plakans, L. (2009). Investigating source use, discourse features, and process in integrated writing tests. In *Spaan fellow working papers in second/foreign language assessment* (Vol. 7, pp. 47–84). Ann Arbor: The University of Michigan.Google Scholar

Gebril, A. (2021). Learning-oriented language assessment: Putting theory into practice. New York: Routledge.

Gebril, A., & Plakans, L. (2009). Investigating source use, discourse features, and process in integrated writing tests. In *Spaan fellow working papers in second/foreign language assessment* (Vol. 7, pp. 47–84). Ann Arbor: The University of Michigan.Google Scholar.

Gholami, J. and Alinasab, M. (2017). Source-Based Tasks in Writing Independent and Integrated Essays. *International Journal of Instruction,* 10(3), 127-142.DOI: 10.12973/iji.2017.1039a.

Graesser, A. C., McNamara, D., Cai, Z., Conley, M., Li, H., & Pennebaker, J. (2014). Coh-Metrix measures text characteristics at multiple levels of language and discourse. *Elementary School Journal*, 115(2), 211-229.

Grant, L., & Ginther, A. (2000). Using computer-tagged linguistic features to describe L2 writing differences. *Journal of Second Language Writing,* 9, 123–145.

Guo, L., Crossley, S., & McNamara, D. S. (2013). Predicting human judgments of essay quality in both integrated and independent second language writing samples: A comparison study. *Assessing Writing,* 18, 218–238.

Kim, E. J. (2017). The TOEFL iBT writing: Korean students' perceptions of the TOEFL iBT writing test. *Assessing Writing,* 33, 1-11.

Liao, L. (2020). A Comparability Study of Text Difficulty and Task Characteristics of Parallel Academic IELTS Reading Tests. *English Language Teaching,* 13(1).

Oxford, R. (2006). "Task-based language teaching and learning: An overview", *The Asian EFL Journal Quarterly*, 8 (3), 94-121.

Plakans, L., Gebril, A., & Bilki, Z. (2019). Shaping a score: The impact of fluency, accuracy, and complexity on integrated skills performances. *Journal of Language testing.* https://doi.org/10.1177/0265532216669537.

Plakans, L., & Gebril, A. (2012). A Close Investigation into Source Use in Integrated Second Language Writing Tasks. *Assessing Writing*, *17*(1), 18-34.

Plakans, L., & Gebril, A. (2017). Exploring the relationship of organization and connection with scores in integrated writing assessment. *Assessing Writing*, 31, 18-34. http://dx.doi.org/10.1016/j.asw.2016.08.005 1075-2935.

Plakans, L., & Gebril, A. (2013). Using multiple texts in an integrated writing assessment: Source text use as a predictor of score. Journal *of Second Language Writing,* 22, 217–230.

Riazi, A.M. (2016). Comparing writing performance in TOEFL-iBT and academic assignments: An exploration of textual features. *Assessing Writing, 28,* 15-27. DOI: 10.1016/j.asw.2016.02.001.

Shi, L. (2004). Textual Borrowing in Second-Language Writing. *Written Communication*, 21, 171-200.

Soleimani,H., and Mahdavipour, M. (2014). The effect of variations in integrated writing tasks and proficiency level on features of written discourse generated by Iranian EFL learners. *Journal of Teaching Language Skills. (JTLS)6 (2),* Ser. 75/4.ISSN: 2008-8191. pp. 131-159.

Stricker, L., & Attali, Y. (2010). Test takers' attitudes about the TOEFL iBT (TOEFL iBT™ Report No. iBT-13). Princeton, NJ: ETS.

Yu, G. (2008). Reading to summarize in English and Chinese: A tale of two languages? *Language Testing*, *25*(4), 521–551.

Weir, C. J., Huizhong, Y., and Yan, J. (2000). *An Empirical Investigation of the Componentiality of L2 Reading in English for Academic Purposes.* UCLES Cambridge: Cambridge University Press.

Weir, C. J. (2005). *Language testing and validation*: An evidence-based approach. Basingstoke: Palgrave Macmillan.

Weir, C., & Shaw, S. (2006). Defining the constructs underpinning main suite writing tests: a socio-cognitive perspective. *Research Notes*, 26(3), 9-14.

Weir, C. J. (1993). *Understanding and developing language tests*. New York: Prentice Hall.

**Biodata**

**Farah Shooraki** is a faculty member at the Department of English, Yazd, and Maybod branches, in Iran. Her research interests include Language Teaching and Learning Teaching, language assessment, and Translation.
Email*: f.shooraki@gmail.com*

**Dr. Hossein Barati** is an associate professor of Applied Linguistics at the University of Isfahan, Iran. His research interests include language assessment, translation studies, Testing, Language Teaching, and Language Learning Strategies.
Email: *barati@fgn.ui.ac.ir*

**Dr. Ahmad Moinzadeh** is an associate professor of Linguistics at the University of Isfahan, Iran. His research interests include Applied Linguistics and syntax.
Email: *Moin@fgn.ui.ac.ir*