



---

## Error Taxonomy of TOEFL iBT Writing: An Iranian Perspective

---

Arshya Keyvanfar<sup>1\*</sup>, Reyhaneh Vafaei Khoshkhou<sup>2</sup>

<sup>1</sup>Islamic Azad University, North Tehran Branch

<sup>2</sup>Islamic Azad University, North Tehran Branch

---

Received: 2 February, 2015

Accepted: 25 August, 2015

---

### Abstract

TOEFL iBT has turned recently heads to the impacts language tests can have on language learning. Since error analysis-based instruction has gained a new life with the advent of the computer analysis of the learner's language, the researchers of this study embarked on examining a sample of integrated and independent writing tasks of 45 Iranian TOEFL iBT candidates in order to identify and classify their errors. Examining 90 writing tasks in toto, the researchers categorized the errors into three major categories: grammar, mechanics, and content errors. Although, the errors of integrated and independent tasks at the levels of grammar and mechanics, with slight differences, had much to share, the content errors exhibited considerable differences. The content errors of the integrated task included plagiarism, for instance while the independent category was limited to organization of ideas and task fulfillment. Given the relative youth of TOEFL iBT for language teachers and learners, the results of this study can have promising potential for enhancing Iranian TOEFL iBT teachers' understanding of the problems of their prospective candidates. The findings can also be used to improve TOEFL iBT preparatory materials as adopting a troubleshooting approach seems to be more viable in short-term tailor-made courses.

**Keywords:** Error Analysis (EA), TOEFL iBT writing, Independent writing tasks, Integrated writing

---

### INTRODUCTION

Writing is undeniably the most difficult task for any language learner to master. The reason lies in the fact that while writing, L2 learners should take many issues into account. They should pay attention to higher level skills such as planning, organizing, and revising as well as the mechanics of the writing such as spelling, punctuation, and capitalization (Richards & Renandya, 2002). The problem becomes far thornier when we have before us new types of tasks, like what is commonly seen in TOEFL iBT writing tasks. Given the recent

commencement of TOEFL iBT, there is a paucity of research in countries like Iran. In fact, TOEFL iBT instructors in this country, including the researchers of this study, have come to a unanimous voice that TOEFL iBT is not yet as understood as it should be in Iran. Overreliance of TOEFL iBT instructors on conventional methods of teaching writing, unfamiliarity with its scoring rubrics, and finally inattention to the most common errors of Iranian TOEFL candidates have made the writing tasks of this test more difficult and sensitive than ever. To our knowledge, the present study in Iran is the first to investigate a sample of integrated and independent writing

---

\*Corresponding Author's Email:  
[arshya.keyvanfar@gmail.com](mailto:arshya.keyvanfar@gmail.com)



tasks of a group of Iranian TOEFL iBT candidates in order to identify and classify their most common errors. Though Error Analysis (EA) fell from grace long ago, the researchers of this study like Erdoğan (2005) hold that only EA with its postmortem examination of the candidates' writing is capable of generating new ideas as to how this self-complicating skill of writing can be taught and learned especially for assessment purposes. It goes without saying that this study primarily aims at demystifying the possible washback of TOEFL iBT on writing instruction in Iran and considers its contribution to the field, if any, as highly ambitious.

### **TOEFL Background and Criticisms**

In 1963, Test of English as a Foreign Language (TOEFL) was developed by the National Council on the Testing of English as a Foreign Language with the aim of testing the language proficiency of the nonnative speakers applying for admission in different institutions in the United States. Since its inception, TOEFL has gone through different evolutionary stages, from the paper-based to the computer-based and finally in 2005, to the internet-based format.

Previously the writing component of the TOEFL contained only one independent task. However, this task was seriously questioned since it did not reflect the exact genre used in real academic settings (Ohkubo, 2009). Hamp-Lyons and Kroll (1997, as cited in Cumming, Grant, Mulcahy-Ernt, & Powers, 2005) have censured then-administered writing component of TOEFL test on the grounds that it by no means assessed the types of writing students realistically have to perform in academic settings. In 2005, Cumming, et al. challenged the educational relevance, authenticity and content validity of the TOEFL test. Their most significant criticism of the TOEFL was that it merely focused on the discrete knowledge about language forms, which can be comfortably taught and has definitely a negative washback on learning and teaching.

To obviate this discrepancy, two prototype integrated writing tasks with a single source text

(a reading passage or a lecture) were initially developed. These tasks, however, for two reasons met with partial success only. According to Cumming, Kantor, Baba, Eouanzoui, Erdosy, and James, M. (2006), overreliance of the test takers on the expression obtained from the reading source text without quotation marks or author acknowledgement along with the long time it took to do the two tasks (Pearlman, 2008) were the main reasons of the abandonment of these tasks. As said by Cumming et al. (2005), there was an attempt in the design of the tasks for the new TOEFL test to take the wanted-to-be-assessed construct into account. Moreover, they asked for the integration of language production modalities with tasks that involve both listening and reading, like what the students need to do in a real English academic task.

Regarding the reasons why previous version of TOEFL needed revision, Zareva (2005) put the validity of the test under question. Traditional threshold TOEFL score, 550 (PBT) or 213 (CBT), were considered as the indication of the readiness of prospective university students to study in an English-medium academic setting (Tannenbaum & Wylie, 2004) but Johnson, Jordan, and Poehner (2005, cited in Zareva, 2005) found evidence of the insufficient writing and oral communication skills of the admitted students, which would consequently impede their full participation in academic programs. According to Darling - Hammond (2000), one of the characteristics of authentic assessment is that it entails the integration of several kinds of skill and knowledge. Facing all these, TOEFL had to move one step forward and develop its new iBT version.

### **TOEFL iBT in Brief**

Including tasks reflecting various academic discourse used in higher education, i.e. universities and colleges, the new TOEFL, as the Educational Testing Service (ETS) explains, presents a new approach to assess English language skills, Zareva (2005) contends. He (2005) explains that the new version of the TOEFL is not only an up-dated version of the previous one, but a test that involves

new components, such as integrated speaking and writing sections which assess the test taker's ability to combine information from more than one source and to communicate about it.

Compared to its predecessor, the writing section of TOEFL iBT includes two writing tasks instead of one and the test takers are required to type their responses rather than being given the choice between writing by hand and typing. The integrated task requires the test takers to read an academic passage, listen to a passage of the same topic, and then within 20 minutes write about how the information in the reading and listening passages are related. The reading and listening prompts act as content input which provides test takers with some vocabularies on a given academic topic as well as the genre conventions based on which they can model their response (Zareva, 2005). Note-taking is allowed during the whole exam.

According to Wall and Horák (2008), the independent task, within 30 minutes, requires the candidates to write an essay that expresses their choice or opinion on a specific issue and to augment their viewpoints with enough explanations and support. The two tasks are scored based on separate rubrics (rating scales) but both are reported based on descriptors for levels 1 to 5 which will be later converted to 1 to 30 on the whole test. The rubric for the integrated task puts more emphasis on the accuracy and adequacy of the information taken from the reading and listening prompts while allocating enough credit to grammatical accuracy. The independent essay is also evaluated based on the quality of the writing in terms of the accuracy of the production, appropriate use of vocabulary, task development, and organization of ideas (Alderson, 2009).

### **TOEFL Washback**

Testing experts (e.g. Popham, 1983, cited in Reynolds, 2010) very much rely on the newly-developed and modified high-stake tests to create positive washback as they are all on the path to become more authentic and direct. Throughout the literature of language testing, there is frequent

emphasis on the washback effects of high-stake exams such as TOEFL on teachers' institutional practices (Messick, 1996; Spratt, 2005). According to Zareva (2005), TOEFL washback directly targets language learners and teachers, material and test developers, and even publishers. Andrews, Fullilove, and Wong (2002) claim that language testing findings do influence language teaching in terms of time allocation and the content. Language testing practitioners have come to understand that a test is a reliable instrument aiming language proficiency for academic purposes only if it has an integrated nature of the use of language skills (Zareva, 2005). Sanchez (2000) provides further support on the integration of language skills. He believes that written questions based on a reading passage encourage learners to read the passage more attentively and clear the misunderstanding of the first reading.

Zhang (2009) in this regard comments that the integration of oral and reading skills involves different strengths of the students while providing interactive opportunities to focus on both receptive and productive language skills. He thinks that this is exactly what the new version of TOEFL has done. As should be abundantly clear by now, TOEFL iBT, with its integrated nature, has turned heads once more to the impacts language tests can have on language teaching. Apart from its undeniable role on the test takers' lives, TOEFL iBT has brought integrated teaching of language skills to the very first of agenda. Expectedly, TOEFL preparatory classes, and even in the larger scope, general English classes should base their teaching on skills integration approach, and if they do so, then according to Reynolds, (2010) TOEFL has achieved one of the main goals of its development, i.e. encouragement of positive washback.

### **Theoretical Background of Error Analysis and Error-Based Instruction**

Over years the concept of error and error related issues have been hotly debated and researched from different viewpoints. Interwoven with Structuralism and Behaviorism, Contrastive Analysis (CA) claimed that the only barrier in learning a second

language is the interference of L1 system with the L2 system, and stressed that a structural and scientific analysis of the two languages would enable people to predict the areas of difficulty and confusion (Fang & Xue-mei, 2007). Behaviorists firmly believed that errors are the result of L1 interference, and they should be treated rigidly since they lead to formation of bad habits. The followers of CA believed that the errors should be corrected right on the spot; otherwise, these errors will lead to the formation of bad habits. Despite its very promising beginning, CA had a very short heyday. Among its glaring shortcomings, Carter and Nunan (2001) refer to three major weaknesses. First, CA remained paralyzed in explaining errors unattributable to the learner's L1. Second, the study of the learners' errors revealed that their errors are systematic rather than random, and finally it appeared that as language learners develop competence in the target language, they move through a series of successive stages each of which are characterized by particular types of errors. To this list, Fang and Xue-mei (2007) add the CA's overemphasis on outer environment which leaves the language learner out of the picture.

Due to the drawbacks of contrastive analysis, Error Analysis (EA) came as an alternative. EA very soon became the prey of the researchers, for it was considered to hold in it the keys to understand the process of second language learning. EA consisted of preparing a sample of L2 learners' errors, sorting these errors into types, and assuming possible sources for the errors (Johnson & Johnson, 1999). In 1959, Chomsky brought a major shift in the perspective. He, from a cognitive point of view, regarded errors as the result of students' thinking through the process of rule formation. Not surprisingly, language educators and linguists turned their attention from CA and focused on the specific language the learners use attempting to communicate in the target language. Introducing "Interlanguage", in 1972, the English linguist, Selinker, regarded errors as a natural step in students' interlanguage development. Interlanguage is "a system that has a struc-

turally intermediate status between the native and target languages" (Brown, 1994, p. 203). Nemser (1971) and Corder (1971) respectively used the terms *Approximate System and Idiosyncratic Dialect* to refer to the notion of Interlanguage (cited in Fang & Xue-mei, 2007). Corder (1967) distinguished learners' errors as a way to obtain knowledge about the processes and strategies of Second Language Acquisition.

Corder, the pioneer of Error Analysis, was the first to look at errors from a new perspective. It was in his article "*The Significance of Learner Errors*" (1967) that he argued against the popular idea of eradicating errors. He further accentuated that learners' errors should be accepted as a part of the language learning process and firmly claimed that errors are important in themselves. Corder (1967) argued that investigation of errors can be both diagnostic and prognostic at the same time. He explained it is diagnostic because it tells about the learner's knowledge of language at a given time during the process of learning, and it is prognostic in the sense that it provides course organizer with a clearer picture of the learners' current problem; therefore, they can reorient the materials based on the learners' needs and weaknesses. However, it was with the emergence of Communicative Language Teaching that errors found a new identity. In this view errors are not important unless they hinder communication

Contributions and Limitations of EA in Second Language Acquisition (SLA)

According to Ellis (2008), Error Analysis has come under fire on both methodological and theoretical counts. Given the importance of acknowledging the learner language as a variety in its own right, EA is seriously questioned for its Comparative Fallacy, as it puts "too much emphasis on the deviation of the learner's interlanguage (L1) system from the TL system" (Bley-Vroman, 1983, cited in Purdy, 2004, p. 1). Later, James (1998) boldly responded this criticism arguing that because learners are targeted on the norms of native speakers, they perform cognitive comparisons in the process of L2 learning. While Doughty (2001) believes that cognitive comparison is a

“cognitive intrusion designed to enable mapping between a conceptual representation and a new linguistic form under the influence of pragmatic, semantic, syntactic, and morphological information” (p. 88), James (1998) views this as a form of error analysis and hence an activity more associated with teachers and researchers not students.

Another charge leveled against EA is that it is mostly cross-sectional in nature, that is, it examines learner language at a single point, so it cannot clarify the developmental stages the learner goes through unless diachronic data are used (Jie, 2008). Ellis (2008), on the contrary, sees no deficiency in EA from this perspective because he believes that EA can be perfectly conducted in a longitudinal fashion, by collecting samples of learners’ language at different points in time. Jie (2008) finds the difficulty of collecting large databases of learners’ language, as the sole reason why no robust error typology covering all language errors has been developed to date in traditional EA. Schachter (1974) refers to another drawback of error analysis, suggesting that there is no single cause for a particular error. She thinks that some errors are due to L1 interference, some are attributable to L2 and others happen under the influence of both L1 and L2. Brown (2000) sees the overemphasis of error analysis on production data as another shortcoming of this line of practice. He holds comprehension along with production can finally give a better understanding of the process of SLA. Fang and Xue-mei (2007) also argue that since error analysis has mainly utilized language learners’ production, it has disregarded the equally important comprehension data in providing evidence of language acquisition process.

In her study, Schachter, like James (1998) and Tarone (1981), further reminds us that error analysis fails to account for the avoidance strategy. She found that the absence of relative clause errors in Japanese speakers in contrast to Persian speakers could not be inferred as nativelike competence. Decades later, Fang and Xue-mei (2007) too point to the error analysis failure in dealing

with avoidance strategy. Meanwhile, they warn against too much attention on the errors which may lead to leaving correct utterances unnoticed. Gass (1989, cited in Brown, 2000) refers to the same problem stating that EA’s preoccupation with specific languages should not deter us from viewing linguistic aspects and elements that all languages share.

However, EA still serves to investigate a specific research question rather than as a means to provide comprehensive picture of the idiosyncratic forms of the language learners (Ellis, 2008). Among the advantages of conducting EA, Xue-mei and Fang (2007) refer to the followings: First, EA gives teachers a general knowledge of the learners’ errors; it is through EA that teachers learn to tolerate the errors because making error is an inevitable stage in language learning. Second, errors are in fact invaluable feedback, since they can tell how much learners have improved toward the goal. Third, errors are indispensable device which learners use to learn. And finally, errors should be attended and handled because they will otherwise be fossilized.

Concerning how teachers can benefit from Error Analysis, Corder (1987) states that errors tell what learners have already learned and what remains for them to learn. As Erdoğan (2005) believes, errors function as feedback in the sense that they reflect whether the teaching style of the teacher is effective and what changes it still needs. Errors, he adds, can identify the points and areas which need further attention. He concludes that any study targeting EA is aimed at investigating the language learners’ strategies, the reason why language learners make errors, and the common difficulties in learning and development of remedial materials.

Ellis (2008) agrees that error analysis plays a major role in remedial approaches to teaching of writing. Following the role of EA in language teaching, Ferris (2002) revived Hendrickson’s (1978) appeal to examine the errors of the language learners as the grounds for deciding what L2 features to teach. Ellis (2008) also talks about the use and function of EA as a measure

of accuracy in studies investigating the impact of task design and implementation on the learner production.

Most importantly, EA has gained a new life with the advent of the computer-based analysis of the learner language. The corpus collected for International Corpus of Learner English (ICLE) program is one of the learner error-tagged corpora (Ellis, 2008). Ellis (2008) suggests that error-based analysis gives the opportunity to conduct EAs in large scopes. Though there are ample studies on the writing errors of EFL learners and different types of feedback on their production, few research studies, if any, has conducted error analysis on high stake exams such as TOEFL.

### **Error Correction and Error Feedback**

Under the influence of Truscott's 1996 article, "The Case against Grammar Correction in L2 Writing Classes", the notion of corrective feedback has won considerable interest of language researchers and classroom practitioners. Truscott (1996) spoke out against the place of grammar correction in writing classes and emphasized its abandonment. Analyzing the studies of Kepner (1991), Semke (1984), and Sheppard (1992), he made the point that no research has introduced convincing evidence showing the benefits of error correction on accuracy improvement of students' writing. He further claimed that error correction is also harmful in that the time and energy that can be spent on advantageous aspects of writing programs are now diverted to error correction in vain. Pointing to the increasing research findings showing effectiveness of error correction in writing classes, in a response to Truscott's negative position (1996), Ferris (1999, cited in Bitchener, Young & Cameron, 2005) introduces Truscott's ideas as premature and extremely strong. Huang (2009) finds both immediate error correction and overcorrection in extremes. He believes too much adherence to error correction would result in learners being scared to make mistakes, which will consequently affect their bravery to communicate

in the target language. Huang (2009, p. 86), however, comments that encouraging communication and speaking in English with no error correction leaves the learner on the other extreme which is the development of "Mute English".

Similarly, in an attempt to answer why correction of errors is needed, Ur (2000) states that students prefer and expect to be corrected simply because they want what they produce in L2 to be understood, and since they are not aware of the errors they make, they need teachers' feedback to focus their attention to their errors. Huang (2009), referring to error-correction theory principles, notifies that language teachers should consider the affective factors of language learners and should try to minimize their impatience toward the errors. He also emphasizes that creating a healthy, relaxed, stress-free classroom environment can help students a lot and the inappropriate error-correction can do harm to students with poor English backbone and of introverted characteristic. Erdoğan (2005) suggests that the technique of error correction is something beyond mere repeated presentation of the data and going through the same drills. He thinks it specifically requires the identification of the source of the error, which will in turn help the teacher decide what type of treatment is needed. It is unfortunate, however, that in many cases, despite all the suggestion on teacher's mild attitude to learners' errors, language teachers often take intolerant and negative attitudes toward learners' errors (Xue-mei & Fang, 2007) causing learners to benefit so narrowly and all the time, energy, and attention allotted to error correction to be wasted.

Considering all these, this study embarked on presenting error taxonomy of the TOEFL iBT in the hope that this taxonomy could help language teachers to better understand the most common types of errors Iranian candidates typically make while attempting the new version of TOEFL writing section.

## Design

### Participants

The participants were 45, 22 male and 23 female, Iranian TOEFL iBT candidates. They were all Persian native speakers ranging in age from 21 to 46 with an average of 29. All the participants had about 2 years of English learning experience and were from different educational and vocational backgrounds. They had all completed high school

diploma requirements and were mainly undergraduates or graduates of engineering and medicine-related fields. The participants' TOEFL writing tasks provided the convenient sample for the error analysis. The sample was clinically collected as it was specifically collected for research purposes. Based on Ellis's (2008) list of factors to be considered when collecting learner language samples, the following table was drawn:

**Table 1**  
*Characteristics of the Learner Language Sample of the Study*

Factors	Variables	Description
<b>Learner</b>	1.Proficiency level	Intermediate, using TOEFL
	2.Language learning experience	English language classes (6-18 months), job related English courses (not exceeding 4 months)
<b>Language Sample</b>	1.Medium	Written
	2.Genre	An integrated writing & an essay
	3.Content	Academic & argumentative
<b>Production</b>	1.Planned	The discourse produced allowed for respectively 20 & 30 min task completion

## Instrumentation

The participants took a mock test from *The Official Guide to the New TOEFL iBT: ETS* (2006). Cronbach alpha reliability calculated for the listening and reading sections were .91 and .89 respectively. Writing and speaking sections also showed interrater reliability of .88 and .86. The test results indicated that the candidates' scores on the whole test were between 60 and 73. The writing tasks of the participants were rated using ETS TOEFL iBT writing rubrics (see Appendices A &B) and the scores came to range between 2 and 3.5 which can be also reported as ranging between 14 and 22 (see Appendix C). It is legitimate to conclude that the participants were intermediate in terms of language proficiency level. The language school where the participants took the test has a policy of placing TOEFL candidates in preparatory courses only if their scores are 50 and above. In this regard, it can be claimed that the participants of the study were typical TOEFL students of this language school.

## Procedure

Throughout this study, the researchers had the aid of a TOEFL iBT teacher and an English native speaker, a 28-year-old Canadian Journalism graduate. The writing tasks of these candidates were scored by the TOEFL iBT instructor and one of the researchers. In this process, the raters examined a total of 90 integrated and independent tasks to spot and categorize the most frequent errors of this sample of Iranian TOEFL candidates. At first, grammatically erroneous sentences of all the participants were identified and were subsequently given to the native colleague to provide their reformulated version. The grammatical errors were then categorized and named based on the grammatical structure violated.

Investigating the writing tasks for any possible error, the researcher and her colleague investigated the writing tasks in terms of the mechanics of the writing and the content. The errors of both integrated and independent tasks were subsequently placed in three major categories of (1) grammatical errors, (2) content errors, and (3) mechanical errors.

Despite some commonalities between the errors in the integrated and independent tasks, the errors of the two writing tasks were categorized separately.

## Results

### *Error of Grammar*

In terms of grammatical errors, the integrated and independent tasks, with slight differences, share a lot of features. Therefore, in the following section, the grammatical errors of both tasks are jointly tabulated and described. Table 2 provides definitions and the examples for each identified error category at the grammatical level. It should be reminded that the last category, connectives, appeared in the independent task only. One possible explanation for this can be attributed to the fact that in integrated task students may have found an example of the connective word and used it correspondingly in their writing.

As Table 3 demonstrates, in total 1134 errors were identified in the integrated and independent tasks. As shown, among the grammatical error categories, verb errors with the frequency of 319 comprise 28% of the errors, and seem to be a real problem for the Iranian candidates at this level. Next, agreement errors with 23%, following verb errors, are the most frequent types of error in both tasks. Part of speech and sentence structure respectively with 14% and 13% are in the third and fourth most frequent positions. The category of article and determiner errors, comprising 10% of the all errors, stands in the fifth place. Preposition errors also include 7% of all errors. The connective errors category, the only independent-task specific error of this set, forms only 4% of the errors. Finally, wrong word category, forming merely 1% of all errors, is at the bottom of this frequency hierarchy.

**Table 2**  
***Grammatical Error Categories of Integrated and Independent Tasks***

<b>Grammatical errors</b>	<b>Definitions</b>	<b>Examples</b>
<b>Verb</b>	Errors in tense, passive/active, infinitive, gerund, modals	1. <i>Team let you to be (be) more creative.</i> 2. <i>Teamwork helps members to understanding (to understand) that...</i>
<b>Agreement</b>	Errors in subject-verb, number, and pronoun agreement	1. <i>There are (is) a group of people who like to work alone.</i> 2. <i>In addition to what I said, I think telling the truth set (sets) you free</i>
<b>Preposition</b>	Any omitted, unnecessary, and incorrect uses of verb & adjective prepositions	1. <i>Some people are not interested to (in) group work.</i> 2. <i>There are (is) a group of people who like to work alone.</i>
<b>Part of speech</b>	Errors of grammatically incorrect part of speech	1. <i>I can success (succeed) more...</i> 2. <i>I should tolerant (tolerate) teamwork.</i> 3. <i>We know that single people can't work good(well).</i>
<b>Sentence structure</b>	Any error in main & subordinating clause structure	1. <i>There are many times which (when) we see...</i> 2. <i>I didn't know (whether, if) I should tell her or not.</i>
<b>Article/determiner</b>	Any omitted, incorrect, or unnecessary use of the article "the" and other determiners	1. <i>Team work can provide (a) wider range of knowledge...</i> 2. <i>They can't come to (a) conclusion...</i>
<b>Wrong word</b>	Incorrect spelling resulting in formation of a meaningful actual English word	1. <i>I lie to my friend because I founded that truth is not good.</i> 2. <i>We can think of some advantageous (advantages) of teamwork.</i>
<b>Connectives</b>	Any omitted, incorrect or unnecessary use of connective words, expressing cause & effect, result	1. <i>She sometimes lies (because of) she is honest.</i> 2. <i>I couldn't tell my mom, (so that) I lied.</i>



**Table 3**  
**Frequencies and Percentages of the Grammatical Errors**

Errors	Task One		Task Two	
	Frequency	Percentage	Frequency	Percentage
Verb	187	33.9	132	22.64
Agreement	113	20.5	148	25.38
Part of speech	69	12.5	90	15.43
Sentence structure	81	14.7	66	11.32
Article/determiner	61	11.07	52	8.91
Preposition	36	6.5	43	7.37
Connectives	Ø	Ø	45	7.71
Wrong word	4	0.7	7	1.2

*Mechanics Errors.* At the level of mechanics, both tasks showed to have more or less the same error types, so this level is discussed for both under the same title of mechanics errors.

The following present the error categories, definitions, and examples of this level.

1. Capitalization errors: All uses of small letters for capital letters and unnecessary uses of capitalization.

- When i work with a team, i am very good worker.
- People like To work with friends.
- When I was smaLL i was a good liar.

2. Punctuation errors: All cases of omission, incorrect use of period, colon, semicolon, omitted and incorrect punctuation in subordinated clauses. All required but missed punctuations are indicated in parentheses.

- When i was small (,) i was a good liar (.)
- i can success more when i am in (a) team(.)

As for the mechanics errors, the letter “T”, interestingly, seemed a matter of confusion. In total there were 82 instances of incorrect capitalization of this letter. There were also 57 instances of small letter “i” meaning the first singular person. However, among these writing productions, some students showed occasional awareness of the difference between “I” meaning the first singular pronoun and the small letter “i”; in other words, they were sporadically able to distinguish the difference between the singular pronoun “I” and the small letter “i” as an alphabetic letter. Capitalization of the letters “T”, “J”, and “L” seems to have a special

place. Regardless of its position in a word or the whole sentence “T” was capitalized for 27 times in 15 writings and letter “J” for 18 times in 9 writing compositions. Punctuation errors were mainly omission of periods at the end of sentences, omission of commas when writing a list of words as well as omission of semicolons. In fact, there were rare, if any, instances of misused punctuation at all; they were mostly omission of required punctuation.

*Content errors.* At the content level, as expected, the integrated and the independent tasks differed considerably. The most obvious reason can be attributed to the different nature and focus of these two tasks. Consequently, at the content level, each task was analyzed and explained separately. Based on the ETS rubrics of integrated and independent tasks (see Appendices A & B), the content of writing productions were investigated to determine if they were mannerly produced and appropriately reflected the content desired.

➤ Content in the integrated task: Investigating the integrated task, the researchers finally came to classify the content problems of the integrated writings into three major categories for each of which definitions and examples are provided below.

1. Plagiarism: It referred to all sentences/phrases taken intact and unaltered from the reading and/or listening passages. The underlined structures present some plagiarized language found in the candidates’ integrated writing production.

- *Team work creates new solutions because a group is more likely to make risky decisions that an individual might not undertake.*
  - *Group work is better because a team's overall results can be more far-reaching and have greater impact.*
  - *Team work is better because a group of people has a wide range of knowledge, expertise, and skills than any single individual. Therefore (,) the group spreads responsibility (.)*
2. Own idea addition: This type of error includes all sentences reflecting personal ideas and conclusions made based on personal understanding rather than what the reading and listening passages reflected. In the following examples the underlined words are the personal ideas of the students.
- *Scientist recently find (found) group work as a way to do some jobs in companies.*
  - *At school, we learned many things about group work but I don't like it.*

*We know that single people can't work good (well) and companies have group people.*

It should be abundantly clear by now that integrated task of TOEFL is mainly judged based on the content of the writing production. In total, 35 from 45 participants had either plagiarized sentences or own idea addition or even both in their writings. Interestingly, while female candidates, 19 out of 23, preferred plagiarism, only 11 out of 22 male participants plagiarized in their writings. Comparatively, male candidates, 16 out of 22 comprising 35% of the candidates, were pioneer in adding their own ideas whereas only 8 out of 23 female participants put their view points in their writing productions. Any non-referred sentence involving exact consecutive words of reading and listening passage was regarded as plagiarism. And any point, not mentioned in the reading and listening passages, regardless of the number of

words used to express it was considered addition of own idea. The candidates differed in the plagiarism and own idea addition in terms of the number of words they used.

3. Question addressing: This type included all cases of inadequately addressing the question. Integrated task questions require the test taker to have a grasp of both reading and listening passages. As far as the researchers have noticed, the questions are mainly in three formats. 1) the question requires summarizing the information of the reading and the listening, 2) the listening passage introduces an idea or example in disagreement to what was stated in the reading, hence casting doubt on the point in the reading, 3) the listening and reading are complementary in that they both support the same view, with the listening usually providing an example and/or an additional reason, point, etc.

The result showed that 46% of our participating candidates, 21 out of 45, only focused on either the lecture or mainly the reading passage information, while integrated task generally demands a combination and understanding of both sources of knowledge; subsequently, they failed to appropriately address the question. There were 9 out of 45 who included the information of both input sources; however, only 4 of them were successful to adequately associate the ideas of both passages as the question demanded.

- *Content in independent task.* Based on the ETS rubrics of the independent tasks, we investigated the independent tasks to determine if the topic was properly elaborated and addressed, and if the candidates have produced a well-organized writing showing sufficient exemplification, explanation, and unity to support their viewpoint. Unlike in the content of the integrated task, where three

types of content error were identified, for the independent task the essays were identified as poor, medium, and strong regarding the task development and the organization of the ideas. From a total of 45, 19 (10 female, 9 male) compositions seemed to present more of the requirements of the independent task and were scored higher.

### Discussion

As it was observed, one error category, connective errors, belonged to the independent task only. As an explanation for this, the researchers assumed that reading and listening passages of the integrated task may have provided the test takers with the correct use of the connective word, and consequently prevented the occurrence of the same error in their integrated writing. Given the frequencies of all grammar errors, it was considered that errors of verb, followed by errors of agreement and part of speech, were the most serious errors of the students. As for the errors of mechanics, they came to present three main sub-categories of punctuation, spelling, and capitalization. Although errors of mechanics in most cases do not hinder the communication of the ideas, the researchers believed that this error category, like others, should be taken care of in language learning classes. In the experimental phase of the study, the researchers wrote down student's explanations on the reasons they ignored mechanics of writing. They claimed that they were too preoccupied with the content, ideas, and grammatical requirements that they forgot to take care of punctuation. Besides, some also reported their unfamiliarity and lack of appropriate training of the fundamentals of punctuation. It is very difficult to claim that adding capital letters in the middle of a word is due to the students' ignorance or lack of knowledge, because based on what the students claimed in the four classes of the second phase, they did this because they found their handwriting more eye-appealing.

The most interesting error category was the content errors of the integrated task including plagiarism, own idea addition, and question ad-

ressing. As expected, most plagiarized parts were taken from the reading passage rather than the listening passage. It can be argued that students could get more from the reading rather than the listening passage, because they had the reading text available through the writing time limit. It can be further suggested that plagiarizing from the listening passage required a good command of listening which Iranian TOEFL candidates seem to lack (Keyvanfar & Rezayee, 2011). But a more important question is why plagiarism happened at all? We should bear in mind that the ABCs of writing in western academic settings are different from what we practice in Iran. In Iran at school level, plagiarism is often ignored, to be forgotten for the students to take care of. At university level, now that the lesson load is heavier, most students find it harder to avoid plagiarism, so plagiarism is once more marginalized, if not totally overlooked. Teachers may also be reluctant to address the issue because they may find plagiarism a western-specific concept and sometimes face-threatening (Ha, 2006). It was also observed that male students, compared to females, were more inclined to add their own ideas to the writing productions. This finding may be due to that fact that male English learners in Iran feel more confident in language classes and express their ideas more comfortably.

As far as the last category, question addressing, is concerned, it was noticed that few students came to realize and correctly decode the question of the integrated task. The researchers in the experimental phase directly asked students how they approached the task question. They reported that they believed the question needed summary information of both passages; some also mentioned that the question included some unknown words and they could not fully understand it.

### Conclusion

For the integrated writing task, TOEFL iBT test requires test takers to summarize and connect the main points of a 200-300 word text and a two-minute word lecture. Cumming (2001) says that

the integrated task of TOEFL is evaluated in terms of language use and its content. In a validating study of the TOEFL integrated task, Ohkubo (2009) directs attention to two important requirements of a good performance: a) acknowledgment of the input text as the source information, and b) reformulation of the source text. He continues referring to lack of any direct and explicit hint to identifying the source text in the *Official Guide to the New TOEFL iBT*. The present study investigated these two important factors of the integrated task.

Though grammatical and mechanical errors are important in their own right, the content errors of the integrated task, comparatively speaking, have been viewed to be more important than other error categories of this task. Looking at the scoring rubrics of the integrated task (Appendix A), one can realize the key role of content in the integrated task. Therefore, the present study suggests TOEFL instructors pay special attention to the content of the writing production and consequently provide more content feedback in their integrated writing classes. In academic settings, any student failing to provide appropriate referencing while using other's words is penalized for plagiarism. This holds true for the integrated task of the new version of TOEFL too. The score rubrics and *TOEFL Tips* warn that "test takers receive a score of zero if all they do is copy words from the reading passage" (Educational Testing Service, 2007, p. 33). Ha (2006) directs our attention that in spite of the fact that Chandrasoma and Pennycook (2004) explicitly warn on the cost of plagiarism, they show little, if no, awareness of the fact that students need to be adequately trained about how to make reference and citation in academic settings. She then finds it unfair to unreasonably expect students to write based on the APA style unless they have received enough training. Overseas students are often required to obtain a predetermined TOEFL or IELTS score in order to study in an English-speaking academic setting; however, she says that those students taking IELTS test are required to write two writing tasks, neither of which is

based on their reading of a text and so they do not receive referencing and citation training techniques while preparing for the test (Ha, 2006). She questions expecting them to be aware of something before being trained and once more emphasizes the key role of plagiarism training in language classrooms. Ohkubo (2009) also writes about the importance of paraphrasing in academic settings.

Based on the result of this study, almost all candidates' writing tasks we investigated had plagiarized to some extent. Specifically speaking, there is a need to increase Iranian TOEFL candidates' awareness of the plagiarism which may creep into their work. Arm in arm with Phane Le Ha (2006) and Ohkubo (2009), this study is to give more evidence of the role of plagiarism avoidance training.

Given all these, this study also fully agrees with Ha (2006) and suggests TOEFL iBT instructors in Iran embark on teaching the techniques and guidelines to avoid plagiarism. It further, in a larger scope, recommends schools and universities in Iran enhance the awareness and sensitivity of both teachers and students towards plagiarism at school level, which may in long run, assist students to perform better not only in TOEFL iBT integrated task but also in other academic settings.

Furthermore, this study has once more brought error analysis in classrooms to hopefully conclude that old concepts such as error analysis can be differently viewed and interpreted when combined with new approaches, skills integration in this case. Besides, it was attempted to show error-based instruction can provide a starting point for teachers to know where to begin giving feedback. Writing compositions are often scored based on the language use and the development of the ideas, for the integrated task, the error analysis phase was to present other errors, content errors, which could make the candidates' scores suffer more.

Through the course of this study, the researchers came across several intriguing questions which seem well worth of answering. A

similar study can be done to see whether integrated-skill instruction has any impact on better performance of TOEFL iBT candidates in the integrated task of Speaking Section. More plagiarism from the reading passage due to its availability during the writing process suggests that its removal may alleviate this problem which of course requires a separate study. Adding one's own ideas which were more prevalent in male participants can lead to a series of studies investigating the possible impact of tester variables such as gender, field of study, and critical thinking on their integrated writing. Finally different types of error treatment can open up a whole new chapter of researches aimed to improve TOEFL iBT prep courses.

## References

- Alderson, J. C. (2009). Test review: Test of English as a Foreign Language™: Internet-based Test (TOEFL iBT®). *Language Testing*, 26(4), 621-631.
- Andrews, S., Fullilove, J. & Wong, Y. (2002). Targeting washback – a case-study. *System*, 30, 207-223.
- Bitchener, J., Young, S., Cameron, D. (2005). The effect of different types of corrective feedback on ESL student writing. *Journal of Second Language Writing*, 14, 191-205.
- Brown, D. H. (1994). *Principles of Language Learning and Teaching*. Englewood Cliffs: Prentice Hall Regents.
- Brown, D. H. (2000). *Principles of language learning and teaching* (4th ed.). New York: Pearson Education.
- Carter, R. & Nunan, D. (Eds.). (2001). *The Cambridge guide to teaching English to speakers of other languages*. UK: Cambridge University Press.
- Corder, S. P. (1967). The significance of learners' errors. *International Review of Applied Linguistics*, 5, 161-170.
- Corder, S.P. (1987). *Error analysis and interlanguage*. UK: Oxford University Press.
- Cumming, A., Grant, L., Mulcahy-Ernt, P., & Powers, D. E. (2005). A teacher-verification study of speaking and writing prototype tasks for a new TOEFL. *TOEFL Monograph Series MS-26*. Princeton, NJ: Educational Testing Service.
- Cumming, A., Kantor, R., Baba, K., Eouanzoui, K., Erdosy, U., & James, M. (2006). *Analysis of discourse features and verification of scoring levels for independent and integrated prototype written tasks for the new TOEFL*. Princeton, NJ: Educational Testing Service.
- Darling-Hammond, L. (2000). Authentic assessment of teaching in context. *Teaching and Teacher Education*, 16, 523-545
- Ellis, R. (2008). *The study of second language acquisition* (2nd ed.). UK: Oxford University Press.
- Erdoğan, V. (2005). Contribution of error analysis to foreign language teaching. *Mersin University Journal of the Faculty of Education*, 1(2), 261-270.
- Fang, X., & Xue-mei, J. (2007). Error analysis and the EFL classroom teaching. *US-China Education Review*, 4(9), Serial No.34, 10-14.
- Ferris, D. R. (2002). *Treatment of error in second language student writing*. Michigan: University of Michigan Press.
- Ha, P.L (2006). Plagiarism and overseas students: Stereotypes again?. *ELT Journal*, 60(1), 76-78.
- Hendrickson, J. M. (1978). Error correction in foreign language teaching: Recent theory, research, and practice. *Modern Language Journal*, 62, 387– 398.
- Huang, Q. (2009). Accuracy & fluency- Inspiration from error-correction of interlanguage theory. *Asian Social Science*, 5(2), 84-86.
- James, C. (1998). *Errors in language learning and use: Exploring error analysis*. New York: Longman.
- Jie, XU. (2008). Error theories and second language acquisition. *US-China Foreign Language*, 6(1), 35-42
- Johnson, H. & Johnson, K. (Eds.). (1999). *Encyclopedic dictionary of applied linguistics*.

- Oxford, UK: Blackwell.
- Kepner, C. G. (1991). An experiment in the relationship of types of written feedback to the development of second-language writing skills. *Modern Language Journal*, 75, 305-313
- Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13, 241- 256.
- Ohkubo, N. (2009). Validating the integrated writing task of TOEFL internet-based test (iBT): Linguistic analysis of test takers' use of input material. *Melbourne Papers in Language Testing*, 14(1), 1-31.
- Pearlman, M. (2008). Finalizing the test blueprint. In C. A. Chapelle, M. K. Enright, & J. M. Jamieson (Eds.), *Building a validity argument for the test of English as foreign language* (pp. 227-258). New York: Routledge.
- Purdy, J. D. (2004). The comparative fallacy in studies of the L2 acquisition of unaccusative verbs. *Teacher College, Columbia University Working Papers in TESOL and Applied Linguistics*, 4(1), 1-3.
- Reynolds, J. (2010). *An exploratory study of TOEFL students as evaluators of washback to the learners*. Published master's thesis, University of Queensland, Australia.
- Richards, J. C., & Renandya, W. A. (Eds.). (2002). *Methodology in language teaching: An anthology of current practice*. UK: Cambridge University Press.
- Sanchez, M. A. (2000). An approach to integration of skills in English teaching. *Didactica (Lenguay literature)*. 12, 21-41
- Semke, H. D. (1984). Effects of the red pen. *Foreign Language Annals*. 17, 195-202
- Sheppard, K. (1992). Two feedback types: Do they make a difference? *RELC Journal*, 23, 103-110.
- Spratt, M. (2005). Washback and the classroom: The implications for teaching and learning of studies of washback from exams. *Language Teaching Research*, 9(1), 5-29.
- Tannenbaum, R. J., & Wylie, E. C. (2004). *Mapping test scores onto the Common European Framework: Setting standards of language proficiency on the Test of English as a Foreign Language (TOEFL), the Test of Spoken English (TSE), the Test of Written English (TWE), and the Test of English for International Communication (TOEIC)*. Princeton, NJ: Educational Testing Service.
- Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language Learning*, 46, 327-369.
- Ur, P. (2000). Teaching grammar: what can we learn from research. *The TESOLANZ Journal*, 8, 14-22
- Wall, D., & Horák, T. (2008). The impact of changes in TOEFL examination on teaching and learning in Central and Eastern Europe: Phase 2, coping with change. TOEFL iBT Research Report
- Zareva, A. (2005). What is new in the New TOEFL-iBT2006 test format? *Electronic Journal of Foreign Language Teaching*, 2, 45-57. Retrieved June 29, 2010, from <http://e-flt.nus.edu.sg/>
- Zhang, Y. (2009). Reading to Speak: Integrating Oral Communication Skills. *English Teaching Forum*, 1, 32-34

**Arshya Keyvanfar** is an assistant professor at Islamic Azad University North Tehran Branch and has been teaching at undergraduate and graduate levels in the areas of English language teaching methodology, testing, and research. She is particularly interested in TBLT and its application in EFL settings.

**Reyhaneh Vafaeikhoshkhou** is an MA holder in TEFL. She is now a TESL Ontario certified ESL instructor and TESL theory instructor. Her main research interest is the investigation of teaching internationally qualified examinations such as IELTS and TOEFL.

## Appendix A


**TOEFL.**
**TOEFL iBT Test—INTEGRATED  
Writing Rubrics**

Score	Task Description
5	A response at this level successfully selects the important information from the lecture and coherently and accurately presents this information in relation to the relevant information presented in the reading. The response is well organized, and occasional language errors that are present do not result in inaccurate or imprecise presentation of content or connections.
4	A response at this level is generally good in selecting the important information from the lecture and in coherently and accurately presenting this information in relation to the relevant information in the reading, but it may have minor omission, inaccuracy, vagueness, or imprecision of some content from the lecture or in connection to points made in the reading. A response is also scored at this level if it has more frequent or noticeable minor language errors, as long as such usage and grammatical structures do not result in anything more than an occasional lapse of clarity or in the connection of ideas.
3	A response at this level contains some important information from the lecture and conveys some relevant connection to the reading, but it is marked by one or more of the following: <ul style="list-style-type: none"> <li>• Although the overall response is definitely oriented to the task, it conveys only vague, global, unclear, or somewhat imprecise connection of the points made in the lecture to points made in the reading.</li> <li>• The response may omit one major key point made in the lecture.</li> <li>• Some key points made in the lecture or the reading, or connections between the two, may be incomplete, inaccurate, or imprecise.</li> <li>• Errors of usage and/or grammar may be more frequent or may result in noticeably vague expressions or obscured meanings in conveying ideas and connections.</li> </ul>
2	A response at this level contains some relevant information from the lecture, but is marked by significant language difficulties or by significant omission or inaccuracy of important ideas from the lecture or in the connections between the lecture and the reading; a response at this level is marked by one or more of the following: <ul style="list-style-type: none"> <li>• The response significantly misrepresents or completely omits the overall connection between the lecture and the reading.</li> <li>• The response significantly omits or significantly misrepresents important points made in the lecture.</li> <li>• The response contains language errors or expressions that largely obscure connections or meaning at key junctures, or that would likely obscure understanding of key ideas for a reader not already familiar with the reading and the lecture.</li> </ul>
1	A response at this level is marked by one or more of the following: <ul style="list-style-type: none"> <li>• The response provides little or no meaningful or relevant coherent content from the lecture.</li> <li>• The language level of the response is so low that it is difficult to derive meaning.</li> </ul>
0	A response at this level merely copies sentences from the reading, rejects the topic or is otherwise not connected to the topic, is written in a foreign language, consists of keystroke characters, or is blank.

## Appendix B



TOEFL

TOEFL iBT Test – INDEPENDENT  
Writing Rubrics

Score	Task Description
5	<p><b>An essay at this level largely accomplishes all of the following:</b></p> <ul style="list-style-type: none"> <li>effectively addresses the topic and task</li> <li>is well organized and well developed, using clearly appropriate explanations, exemplifications, and/or details</li> <li>displays unity, progression, and coherence</li> <li>displays consistent facility in the use of language, demonstrating syntactic variety, appropriate word choice, and idiomaticity, though it may have minor lexical or grammatical errors</li> </ul>
4	<p><b>An essay at this level largely accomplishes all of the following:</b></p> <ul style="list-style-type: none"> <li>addresses the topic and task well, though some points may not be fully elaborated</li> <li>is generally well organized and well developed, using appropriate and sufficient explanations, exemplifications, and/or details</li> <li>displays unity, progression, and coherence, though it may contain occasional redundancy, digression, or unclear connections</li> <li>displays facility in the use of language, demonstrating syntactic variety and range of vocabulary, though it will probably have occasional noticeable minor errors in structure, word form, or use of idiomatic language that do not interfere with meaning</li> </ul>
3	<p><b>An essay at this level is marked by one or more of the following:</b></p> <ul style="list-style-type: none"> <li>addresses the topic and task using somewhat developed explanations, exemplifications, and/or details</li> <li>displays unity, progression, and coherence, though connection of ideas may be occasionally obscured</li> <li>may demonstrate inconsistent facility in sentence formation and word choice that may result in lack of clarity and occasionally obscure meaning</li> <li>may display accurate but limited range of syntactic structures and vocabulary</li> </ul>
2	<p><b>An essay at this level may reveal one or more of the following weaknesses:</b></p> <ul style="list-style-type: none"> <li>limited development in response to the topic and task</li> <li>inadequate organization or connection of ideas</li> <li>inappropriate or insufficient exemplifications, explanations, or details to support or illustrate generalizations in response to the task</li> <li>a noticeably inappropriate choice of words or word forms</li> <li>an accumulation of errors in sentence structure and/or usage</li> </ul>
1	<p><b>An essay at this level is seriously flawed by one or more of the following weaknesses:</b></p> <ul style="list-style-type: none"> <li>serious disorganization or underdevelopment</li> <li>little or no detail, or irrelevant specifics, or questionable responsiveness to the task</li> <li>serious and frequent errors in sentence structure or usage</li> </ul>
0	<p><b>An essay at this level</b> merely copies words from the topic, rejects the topic, or is otherwise not connected to the topic, is written in a foreign language, consists of keystroke characters, or is blank.</p>

Copyright © 2005 by Educational Testing Service. All rights reserved.



**Appendix C**



**Converting Rubric Scores to Scaled Scores**  
**Writing and Speaking Sections of the New TOEFL iBT Test**

Writing Rubric Mean	Scaled Score
5.00	30
4.75	29
4.50	28
4.25	27
4.00	25
3.75	24
3.50	22
3.25	21
3.00	20
2.75	18
2.50	17
2.25	15
2.00	14
1.75	12
1.50	11
1.25	10
1.00	8
	7
	5
	4
	0

Speaking Rubric Mean	Scaled Score
4.00	30
3.83	29
3.66	28
3.50	27
3.33	26
3.16	24
3.00	23
2.83	22
2.66	20
2.50	19
2.33	18
2.16	17
2.00	15
1.83	14
1.66	13
1.50	11
1.33	10
1.16	9
1.00	8
	6
	5
	4
	3
	1
	0

Performance on the Speaking and Writing sections of the next generation TOEFL test is evaluated based on scoring rubrics of 0-5 for each of the two Writing items and 0-4 for each of the 6 Speaking items. The tables show how the mean rubric score of the two writing items and the mean rubric score of 6 speaking items are converted to a scaled score of 0-30.