
Consistency of Measured Accuracy in Grammar Knowledge Tests and Writing: TOEFL PBT

SAEIDEH AHANGARI & ALI HAMED BARGHI

Department of English Language, Tabriz Branch, Islamic Azad University – Tabriz, Iran

Bio Data:

Saeideh Ahangari has a PHD in TEFL. She is an assistant professor at Islamic Azad University, Tabriz branch. Her main interests are CALL, Language testing and Task-based language teaching. She has presented and published many articles in international conferences and journals.

Ali Hamed Barghi is an MA candidate. He has been teaching English language and TOEFL preparation courses at language institutes for many years.

Abstract

Almost no language test is void of grammar items, and the reason is probably the assumption that there exists a positive correlation between examinees' grammar knowledge and the actual demonstrable level of accuracy in communication. Meanwhile, examples abound where many examinees relatively do well on grammar knowledge tests despite failing to retain accuracy in real-time communicative activities like writing compositions. This is in line with Widdowson (1990) that during communication, attention is given to communicative economy and meaning, therefore, performance is not supported by an underlying competence. This study focuses on the validity of widely-used grammar knowledge tests as an indicator of examinees' level of accuracy in writing by a many test users. A group of Iranian intermediate English language learners' knowledge of certain shortlisted grammatical points at the end of a preparatory course for the TOEFL was checked by analyzing the results of a discrete-point grammar test, like the 'Structure and Written Expression' section of the TOEFL PBT, and once more by studying their writing samples for the 'Test of Written English' of the same test, as an example of grammar knowledge in use; eventually, the obtained results were juxtaposed for consistency comparison, the results of which suggest that an individual's demonstrable knowledge of certain language forms in a grammar test necessarily cannot be generalized to their accuracy in writing; hence, a certain score obtained even in a valid grammar knowledge test is not necessarily a valid indicator of an examinee's level of accuracy in written discourse.

Keywords: Grammar Knowledge, Accuracy, Writing, Testing

Introduction

Accuracy is the ability to use the language correctly, and grammar instruction in any language teaching/learning program mainly aims at uplifting accuracy in learners for better communication. Buck, Byrnes, and Thompson (1989, cited in Hadley, 2003) refer to accuracy as “the acceptability, quality and precision of the message conveyed” (p. 17). Brown (2001) indicates that accuracy means being “clear, articulate, grammatically and phonologically correct,” while fluency connotes being “flowing and natural” (p.268). And, according to Richards and Schmidt (2002), “accuracy is the ability to produce grammatically-correct sentences, but it may not include the ability to speak or write fluently” (p.204). Teachers who concentrate on accuracy help their students produce grammatically correct written and spoken English. Grammar describes the ways in which words are combined to form meaningful and acceptable sentences, and it consists of semantics, word meanings and their relationships, syntax - how we group and order words to form phrases, clauses, and sentences - and morphology - how words are changed according to their use in phrases, clauses, and sentences.

For linguists, grammar is important because it is the language that makes it possible to talk about language, but for many language teachers and learners, the importance of grammar is associated with the accurate use of language for effective communication. Brumfit (1980) believes that maximally effective communication is the result of working on accuracy. Siti Rohani (2007) also confirms this by mentioning that for a student to communicate effectively as stipulated in the aim of the English Language Program, he or she needs to attain both fluency in the language and grammatical accuracy. It is the role of accuracy in effective communication, therefore, that has made grammar instruction essential to any language teaching/learning practice. Ellis (1989, as cited in Adnan (ND)) asserts that formal grammar teaching can improve learners’ accuracy. And Widdowson (1990) depicts grammar as a liberating force.

Most of us are familiar with the phenomenon of students who know the rules of grammar, but who are, nonetheless, unable to ask for simple directions. According to Larsen-Freeman (2001), nowadays many ESL practitioners view grammar less as a body of knowledge to be studied than as a skill to be practiced and developed. Grammar knowledge is important, but only insofar as it enables students to communicate “accurately, meaningfully, and appropriately” When we focus on testing grammar, Purpura (2004) asserts that the relationship between grammatical knowledge and performance is strong, which suggests that learners showing a high grammatical competence are anticipated to outperform in communicative settings. Farhady *et al.* (2006) indicate that structure, or grammar is the most popular language component in language testing because it permeates all language skills, it is easier than other components to test, and most experts agree on what must be included in structure tests. Moreover, grammar tests attempt to gauge the examinee’s knowledge of accurate forms and arrangements of words, phrases, and sentences.

Grammar knowledge is the most elaborate component of linguistic competence, and Ellis (2008) argues that language examiners cannot ignore linguistic competence (and indeed have not done so). What learners can do with language is to a very considerable extent dependent on what language they know. Rimmer (2006) indicates that grammar is central to language description and a posteriori construct validation of language tests consistently identifies grammar as a significant factor in differentiating between score levels and characterizing overall proficiency. Knowing the importance of grammar, many learners practice language forms out of context to do well on language tests, and many of them gain a segmented partial knowledge of decontextualized language structures as a result of their practice, though, this knowledge may not be functional in language use. As Richards and Renandya (2002) point out, "many students may know a lot of grammar, but they are unable to use their knowledge for any practical communicative purposes" (p.146). According to Swan (ND) as cited in Richards and Renandya (2002), "students do not learn English: They learn grammar at the expense of other things that matter as much or more. They know the main rules, can pass tests and may have the illusion that they know the language, but when it comes to using the language in practice, they discover they lack vital elements" (p.151). Similarly, Decarrico and Larsen-Freeman as cited in Schmitt (2002) assert that "learners will be able to complete exercises satisfactorily when their attention is focused on the grammar, but when their attention shifts to a more communicative interaction, the grammar will be forgotten" (p.29). Jackson (2008) says some evidence suggests that even if highly proficient L2 speakers demonstrate explicit knowledge of particular L2 structures, they might have difficulty utilizing this information during online processing.

Widdowson (1990) believes that during communication, attention is given to communicative economy and meaning, therefore, performance is not supported by an underlying competence. Larsen-Freeman as cited in Celce-Murcia (2001) warns that "if they (students) knew all the rules that had ever been written about English but were not able to apply them, we would not be doing our jobs as teachers" (p.255). According to Rafajlovičová (2010), the testing of grammar continues to feature as a component of many school examinations, commercially available proficiency tests and class-based assessment, where the common practice is to test grammatical competence through decontextualized, isolated sentence formats and discrete-point items, which still seems to hold firm in the testing practice. Kathleen and Kitao (1996) argue that while the testing of grammatical knowledge is limited, it does not necessarily indicate whether the examinees can use the grammatical knowledge in a communicative situation.

Bussmann (1998), in *Routledge Dictionary of Language and Linguistics*, defines proficiency as the ability of functioning competently in one's native or second language, involving a sense for appropriate linguistic behavior in different situations. Richards and Schmidt (2002) define Language Proficiency as the degree of skill with which a person can use a language, such as how well a person can read, write, speak, or understand language.

What language proficiency measures focus upon is to determine the extent of examinees' ability to utilize language at the time of examination. A language proficiency test measures an individual's level of proficiency in his/her native or a

foreign language. TOEFL is a good example of an English language proficiency test. According to Wikipedia, the online encyclopedia, the Test of English as a Foreign Language, or TOEFL, evaluates the ability of an individual to use and understand English in an academic setting. It sometimes is an admission requirement for non-native English speakers at many English-speaking colleges and universities.

According to ETS official website, the 'Structure and Written Expression' (SWE) section contains sentences that test examinees' knowledge of important structural and grammatical elements of standard written English.

Test of Written English (TWE) is a subtest of the TOEFL, which was developed to measure the ability of EFL/ESL students to produce an organized essay or report. This section of the test is used primarily to evaluate the English proficiency of non-native speakers who wish to study in colleges or universities in English-speaking countries.

Inspired by the above mentioned theoretical views about the grammar knowledge, in this study the researcher aims at investigating whether Iranian English language learners can retain the accuracy they are often able to demonstrate in grammar knowledge tests in writing compositions, as an instance of language knowledge in use, or not. This study tries to answer the following question:

Is accuracy a consistent feature in Iranian English language learners' writing performance and their grammar knowledge measured by the Test of English as a Foreign Language – Paper-based Test (TOEFL PBT)?

Method

Participants

The research context was an English language school with coed TOEFL PBT preparatory classes whose students' proficiency level was intermediate and above. Out of 43 applicants who took a placement test, 32 participants, aged 19 to 33, were chosen for the study based on their placement test scores, and they were placed in 2 TOEFL PBT preparatory classes.

Instruments

All applicants took a placement test two weeks prior to the course commencement. 'Oxford University Press and University of Cambridge Local Examinations Syndicate Quick Placement Test Version 2' was administered as a written placement test, which was followed by an interview. *'The Complete Guide to the TOEFL Test, PBT Edition by Cambridge'* was introduced as the main course book. The supplementary books were '504 Absolutely Essential Words 5th edition by Barons' to increase the participants' vocabulary knowledge. The book 'Longman Preparation Course for the TOEFL Test: The paper test by Deborah Philips', was used as the main source for the posttest items.

Design

Since there was an attempt in this study to compare the consistency of the results of two different methods of testing accuracy, the design of the study differed from that of a true experimental research, i.e. there were no control or experimental groups; the dependent variables were the obtained data from the given test, and the

independent variables were the SWE and TWE as two tests measuring accuracy. The study was quasi-experimental; participants were tutored for TOEFL PBT throughout the course, and a course accomplishment test, similar to that of a real TOEFL PBT, was given to the participants.

Procedure

One of the present researchers, who was personally in charge of supervising TOEFL preparatory course's placement test and tutoring the classes, made sure that all of the applicants for the course met the proficiency level requirement for entering the class. Out of 43 participants, those who scored below the required entry mark - 11 applicants - were asked to participate in pre-TOEFL classes first in order to gain the minimum TOEFL entry proficiency level to get the most of the target course. Those who were placed at intermediate level had been chosen based on Oxford placement test. So totally 32 participants attended the class for 20 weeks. There were 60 sessions, each session 90 minutes, 3 days a week. The Complete Guide to the TOEFL Test, PBT Edition by Cambridge was covered as the main course book and all its sections were covered. After the course began, each session was allocated to one of the sections of the test that participants needed to get familiar with, namely: Listening, Structure and Written Expression (SWE), Reading, and Test of Written English (TWE), i.e., each of the four sections above was tutored for 15 sessions along the course, but not in a row. Grammar needed to be taught in SWE sessions explicitly and students' L1 was occasionally used for clarification, if required. The syllabus for the SWE grammar instruction was in accordance with what the course book offered. One reason for focusing on one section of the test only in each session was to practice different skills in a parallel fashion; which definitely did not make the participants feel bored along the course. Furthermore, the plan made it possible for the students to have 3 sessions of interval with focus on other skills and sub-skills before taking a new lesson regarding any skill, say SWE or TWE, which meant that participants had enough time to review their lessons, attend to their homework assignments and get the answer to their questions.

In listening sessions, the participants basically listened to and answered sample listening items of TOEFL PBT, then the answers were checked. The participants exchanged opinions about their answers, and when required, the listening questions were played again, and new words and expressions were highlighted and taught overtly.

In reading sessions, participants were asked to individually read sample passages and answer the following questions. Next, the answers were checked in the class and the students supported their responses. When necessary, the whole reading passage or parts of it was read aloud and complementary explanations about it or the meaning of certain words and/or expressions were provided. The participants, apart from learning a lot of words from reading passages, which proved quite beneficiary in writing compositions, exercised reading comprehension skills and strategies as well.

SWE sessions began with teaching two or three grammar points and participants later answered questions which were directly relevant to the instructed points. Often, extra explanation was needed to be provided while checking answers,

therefore, complementary remarks were discussed while and/or after checking most of the items. The participants not only worked to find the problematic structure in each item of the SWE section as directed in the test, but also they were encouraged to correct mistakes, which was a way of checking their overall understanding of English grammar rules in detail. Here, the researchers had the opportunity to assess participants' knowledge of grammar. At times, L1 was used for clarifying vague points, whose aim was to ensure that participants thoroughly got the grammar lesson. The researchers, however, did not contrast L2 grammar rules with those of L1 while providing explanation. L1 was only an occasional medium of instruction to clarify otherwise problematic topics. The researchers' assumption was that a contrastive comparison of certain L1 and L2 structures would interfere with the authenticity and spontaneity of those would-be errors participants might have had in the TWE.

The participants attended the TOEFL PBT preparatory course for 60 sessions as planned. The course was finished with a test similar to one that participants had taken early in the course onset. The result of this test was of utmost importance to the researchers because of the purpose of study. The participants were given an overall score, however, the two SWE and TWE sections of each examinee's paper were carefully studied for accuracy issues. There were 40 items in the SWE part covering the majority of grammar points participants had learned. For the TWE, examinees had been asked to write a well-organized composition on a given topic indicating that a successful writing would be 300 to 350 words in length. These samples were scanned for grammatical mistakes and the number of inaccuracies for each individual was recorded. Besides, a list of the most frequently seen ungrammaticalities was prepared. In this phase, the available SWE and TWE scores could be juxtaposed for consistency analysis since the participants had all received the instruction they needed in order to be accurate language users. Those SWE items that corresponded to the categorized grammatical mistakes in the writing samples were only chosen to be compared with their equivalents in TWE samples. Paired T-test was employed to compare the number of grammatical mistakes observed and recorded in the overlapping sections of the SWE and TWE for analysis since both set of obtained data, i.e. number of errors in each test section, belonged to a single group of participants.

Results

The two tests administered to measure the participants' accuracy were not alike. The SWE section was basically a discrete-point test of grammar knowledge in multiple-choice format, covering a variety of items; however, the TWE was an integrated form of test, where necessarily not all of the grammatical points tested in the SWE section were available in it for measurement. The researchers needed to establish a common base for the comparison first. Therefore, the TWE samples were carefully studied, and a record of the most frequently seen errors was made ready. In this phase of the study, it was only the accuracy of the sentences in each composition that the researchers paid attention to while reading them. Next, the observed ungrammaticalities were shortlisted into 20 categories as shown in Table 1.

An example for each inaccuracy has been provided. There were obviously other mistakes in compositions too, where, for instance, an individual had forgotten to use third person singular subject with the singular form of a verb, but the same grammar rule had been observed elsewhere in the same writing. Such minor inaccuracies have not been considered for the purpose of the study.

Table 1
Observed ungrammaticalities in compositions

Grammatical Category	Ungrammaticality	Example
1- Present Perfect Tense	Past-time Adverbial Expressions	* I have done it on Friday.
2- Adverbial Clauses of Contrast	Use of the Coordinator 'but'	* Although I was tired, but I continued working.
3- Inversion	Normal Word Order	* Not only they don't help, but also they make problems.
4- Conditional Sentences	Verb Patterns	* I would be glad if you help me with the problem.
5- Coordination	Unparalleled Structures	* A dam stops the flow of water, creating a reservoir and raise the level of water
6- Relative Clauses	Relative Pronouns	* The man which gave me the application form was the manager.
7- Relative Clauses	Repeated Object	* The man that she loved him died in an accident.
8- Modals	Past Modals	* I must studied better.
9- Modals – May and Maybe	Using May be instead	* I may be apply for another job.
10- Adjectives	Order	* An old little man was sitting there.
11- Adjectives and Adverbs	Misuse	* He was speaking so loud that everyone heard it.
12- Adjectives	Comparative / Superlative Misuse	* The third group has the larger population.
13- Definite Article	General Countable Nouns	* The computers are going to replace books.

14- Subjunctive	Using modals	* I suggest that they should consult with other companies first.
15- Passive sentences	Perfect Continuous and Future Continues Tenses	* The program will be being checked again.
16- Reported Speech	Tense / Subject	* He told me I was travelling.
17- Reported Speech	Questions	* They know where is it.
18- Causative Sentences	Passive	* My car checked by mechanic before travelling.
19- Verbs	Prepositions	* He was explaining about the situation.
20- Agreement	Subject-Verb	* Each of them need a class of their own.

After errors had been categorized, corresponding SWE questions were highlighted for comparison. Table 2 shows the corresponding items in the SWE section of the test to the categories in Table 1.

Table 2

Corresponding SWE questions to the listed ungrammaticalities in compositions

Category	Corresponding SWE item
Adverbial Clauses of Contrast	Even though she <u>had been warned</u> not <u>to contact</u> the police, <u>but</u> Karen revealed everything <u>at</u> the police station. A B C D
Inversion	Never for one moment <u>she had</u> imagined <u>that</u> the same <u>calamities</u> <u>could</u> happen to her. A B C D
Definite Article	<u>Schools</u> ought to encourage <u>the students</u> <u>to do</u> creative things <u>besides</u> taking subjects like science and math. A B C D
Conditional Sentences	The marketing specialists believe <u>that</u> if the company had <u>invested in</u> that project <u>two years ago</u> , now it <u>would have made</u> a lot of money. A B C D
Coordination	The electric eel uses its electric shock to capture food and _____. a) for protection b) protect itself c) protecting itself d) it protects itself
Relative Clauses (1)	Somerset Maugham, a novelist, _____ about restless man's quest for inner understanding in The Razor's Edge was a great writer. a) whom he wrote b) who wrote c) wrote d) that wrote
Relative Clauses (2)	Nimbostratus clouds are thick, dark gray clouds _____ forebode rain. a) what b) which c) which they d) what they

Adjectives	The <u>larger</u> of the forty-eight <u>continental</u> states <u>in the</u> United States <u>is</u> Texas. A B C D
Agreement	On <u>the rim</u> of the Kilauea volcano in <u>Hawaiian Islands</u> <u>are</u> a hotel <u>called</u> the A B C D Volcano Hotel.
Verbs	<u>Of the two</u> Diamond islands, <u>only</u> one <u>belongs the</u> United States. A B C D

10 items as shown in Table 2 from among 40 items of the SWE section corresponded to recorded TWE errors; hence, the comparison was made between these 10 categories. Table 3 shows the total number of inaccuracies in the overlapping sections of the two tests.

Table 3

Number of inaccuracies in TWE and SWE sections

Problem Category	Total number of inaccuracies in	
	TWE (32 passages)	SWE (32 tests)
Adverbial Clauses of Contrast	13	5
Inversion	11	9
Definite Article	15	10
Conditional Sentences	6	4
Coordination	12	5
Relative Clauses	6	3
Relative Clauses	3	2
Adjectives	10	6
Agreement	13	4
Verbs	16	14

Total number of inaccuracies for each problem category in the TWE consists of the frequency of the observed problem in 32 writing samples, and the same in the SWE section is limited to the 10 selected grammar items in 32 participants' multiple-choice type tests. Figure 1 shows this frequency in each section.

Figure 1. Inaccuracy types in TWE and SWE section

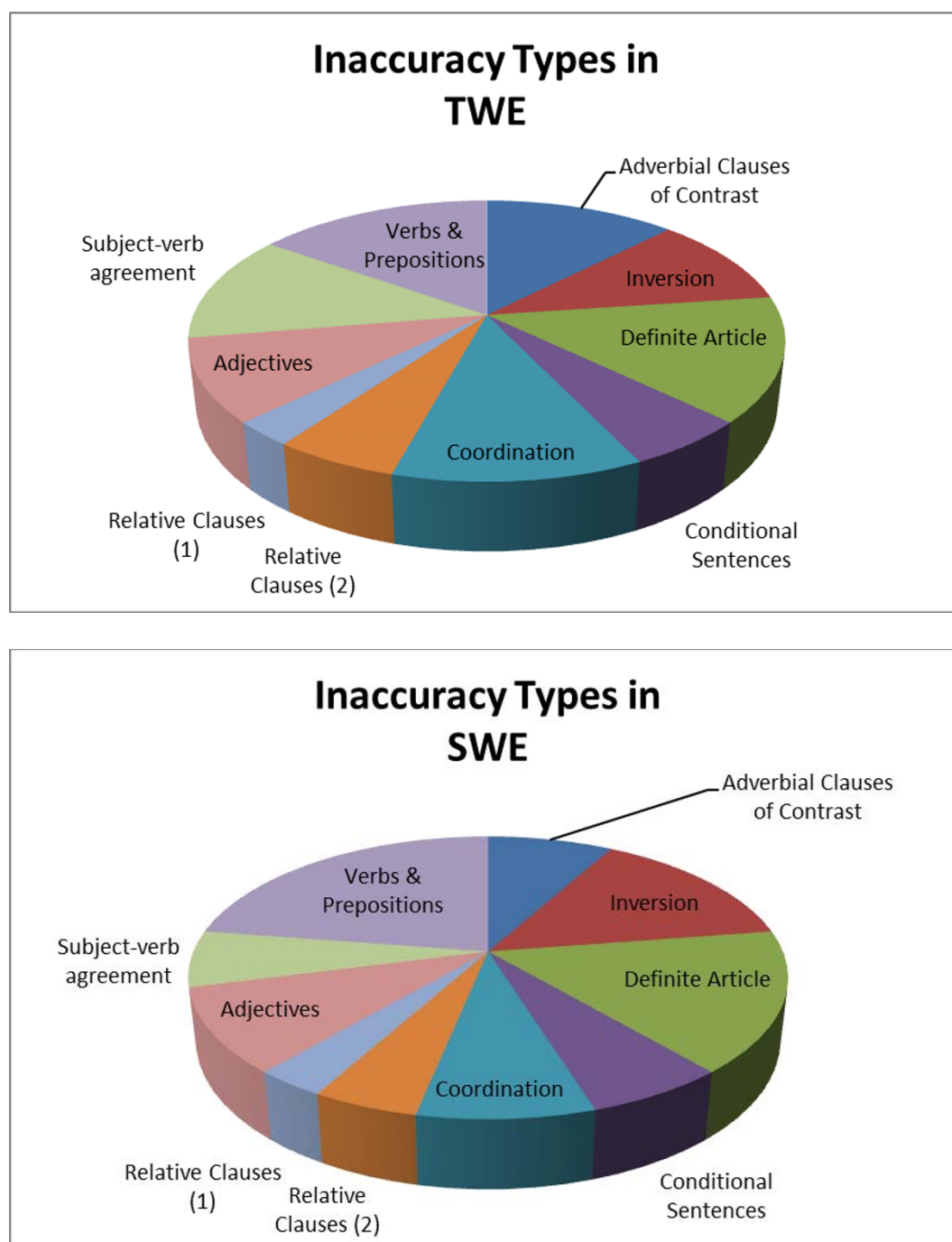


Figure 2 compares the number of observed inaccuracies in the overlapping section of the two tests. A closer look at this figure reveals that the frequency of inaccuracies in the TWE section is considerably higher than the SWE part. The blue bars in the diagram represent ungrammaticalities observed in the writing samples while the red bars are indicative of corresponding error frequencies in the test of grammar knowledge or SWE. The numbers of the horizontal axis refer to the number of each grammar category in table 2 and the figures on the vertical axis are indicative of the frequency of the occurrence of each error.

Figure 2. Juxtaposed TWE and SWE inaccuracies

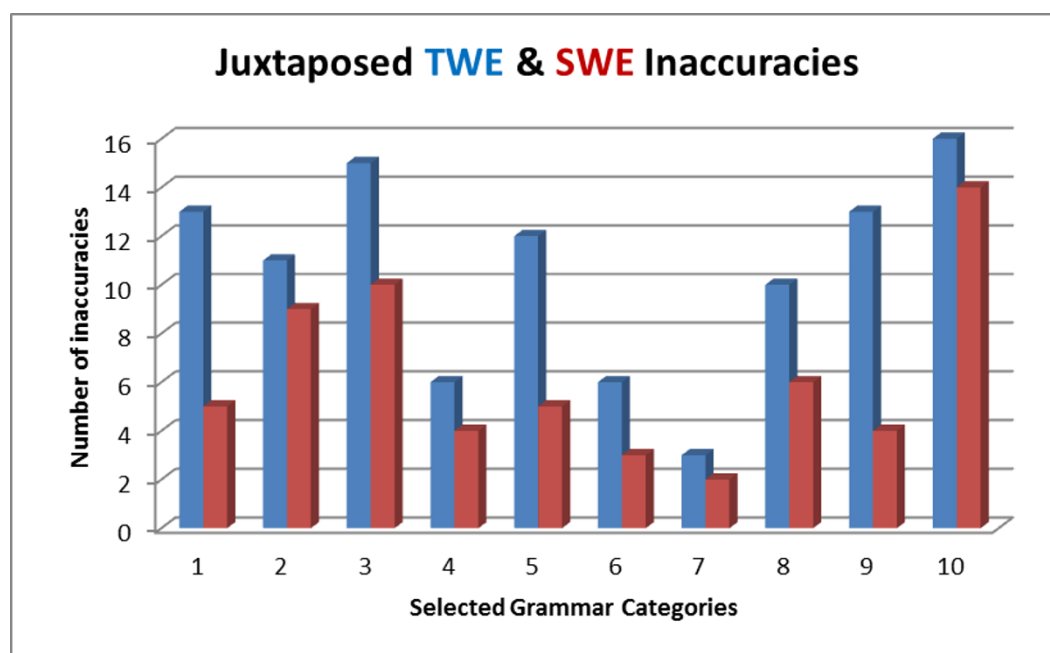


Table 4 shows how poorly each participant has answered the selected items of the SWE section and how many ungrammaticalities he/she had in the TWE. Accordingly, Figure 3 depicts the same, where the horizontal axis is indicative of each participant and the vertical axis shows the number of inaccuracies.

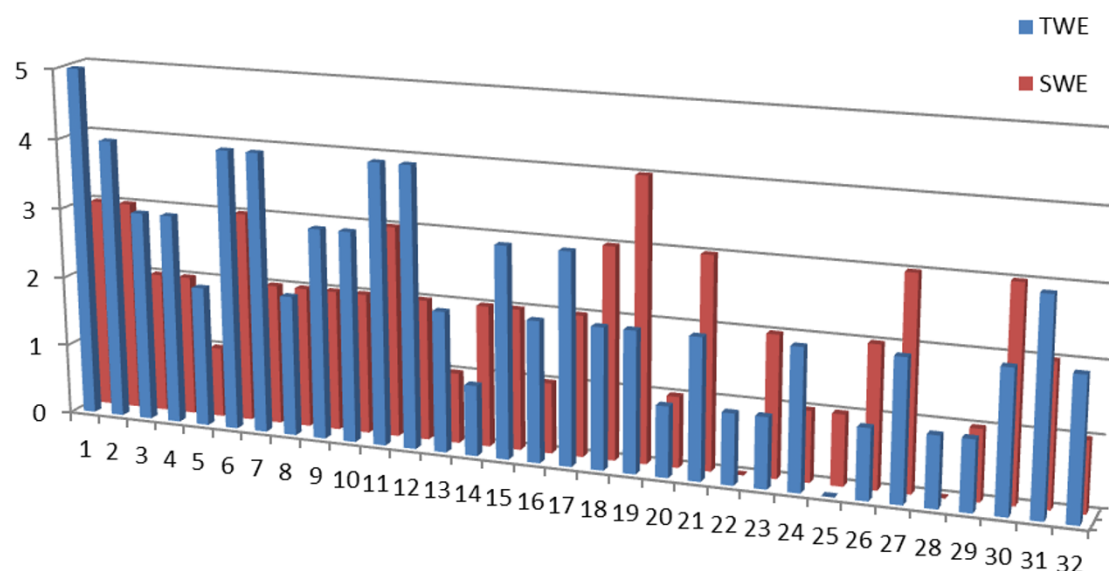
Table 4

Juxtaposed incorrect answers

Juxtaposed Results			Juxtaposed Results		
Participant	Incorrect SWE answers out of 10 selected items	Total number of incorrect sentences in TWE	Participant	Incorrect SWE answers out of 10 selected items	Total number of incorrect sentences in TWE
1	3	5	17	2	3
2	3	4	18	3	2
3	2	3	19	4	2
4	2	3	20	1	1
5	1	2	21	3	2
6	3	4	22	0	1
7	2	2	23	2	1
8	2	3	24	1	2
9	2	3	25	1	0
10	2	4	26	2	1
11	3	4	27	3	2
12	2	4	28	0	1
13	1	2	29	1	1
14	2	1	30	3	2

15	2	3	31	2	3
16	1	2	32	1	2

Figure 3. Juxtaposed incorrect answers



The blue bars represent the frequency of errors in the TWE while the red bars are indicative of the same in the SWE section. A closer look at the diagram shows some exceptions where certain participants have done fairly better in the TWE compared with the SWE section; however, the majority of the participants have had a better performance in the grammar knowledge test or the SWE test. Figure 3 represents the collected data only. In order to make generalizations about the overall performance of examinees the researchers used Paired T-test, whose results have been provided accordingly. Paired T-test was employed for analyzing the differences between the two means of the two tests, based on whose results any consistency between the results of the SWE and TWE with the above-shown set of scores was explored. Table 4 shows the result of this test.

Table 5
Paired T Test results

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	SWE	1.94	32	.948	.168
	TWE	2.34	32	1.181	.209

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	SWE & TWE	32	.510	.003

Paired Samples Test

		Paired Differences				t	df	Sig. (2-tailed)	
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	SWE - TWE	-.406	1.073	.190	-.793	-.019	-2.141	31	.040

The 'Paired Samples Statistics' section of Table 4 shows the mean of the number of inaccuracies observed among 10 corresponding grammatical categories. The mean of errors in the SWE section is 1.94 while it is 2.34 in TWE. It means that the participants, on the whole, had fewer mistakes in the SWE section than TWE; in other words, they had a better performance in the grammar knowledge test than they did in the writing. The 'Paired Samples Test' section of Table 4 shows that the significance value (Sing. 2-tailed) is 0.04, ($t_{(31)} = 2.141$, $p=0.04$), which is less than 0.05 (P value). The difference between the SWE test and TWE results is, therefore, significant. To put it in other words, accuracy is not a consistent feature in the two tests measuring it. Participants relatively did better on the grammar knowledge test than on writing.

It is assumed that the participants, who relatively had done well on the grammar knowledge test, were rather test-wise than grammatically competent in that they had completed a considerable number of grammar test assignments during the course, which were similar to those items they answered in the posttest. The researchers observed that after a couple of sample SWE questions practiced during the course, students could answer similar questions more easily, even if they did not know the meaning of certain words or expressions in the items. It is as if the participants gained a sort of test-wisness as a result of exposure to certain types of grammar knowledge questions and the following explanations provided by the instructor backing them up. So, when the examinee reads a grammar item in the SWE containing a present perfect verb tense, the point that no adverbial expression refereeing to past should accompany it is probably highlighted in his/her mind. However, when the focus is not only on grammar, but also on conveying meaning through concocting sentences, choosing right words and expressions, organizing ideas, attending to spelling, punctuation, etc. the examinee cannot fully utilize his knowledge of grammar, and lapses often occur.

Discussion

Almost all language teachers teach grammar regardless of their variant approaches towards teaching it implicitly or explicitly, so naturally they test grammar as an important language component. 'How to teach grammar effectively' has long ago replaced 'whether or not we should teach grammar'. When testing grammar accordingly, the focus needs to be on the how of doing it; i.e. 'Should grammar knowledge be tested like vocabulary knowledge?' 'Should it be tested in isolation as in grammar knowledge tests?' 'Should it be assessed in language use, instead?' If so, will a composition and/or interview really show how grammatically competent the examinees are? 'How grammatically competent should the one scoring compositions and interviews need to be to judge the accuracy of examinees?'

The findings in this study tend to support Widdowson (1990), who asserts that while someone is engaged in a communicative activity, performance is not supported by an underlying competence because attention is given to communicative economy and meaning.

The researchers also think that the reason why examinees did much better on the grammar knowledge test compared with composition writing, as far as accuracy is concerned, is that writing in its being a communicative activity demands a focus on transferring meaning in a limited time; therefore, accuracy is not the only things to which examinees attend. It is believed that developing ideas to support one's opinion in writing compositions is considered to be more important. Of course, if time limitation allows, examinees might be able to review their written text and check its grammaticality; however, this is naturally done when the idea development stage is over, and there exists a rough content to be revised; besides, not everyone is prompt enough to brainstorm, write a composition and then read it for a second time to check if there are grammatical mistakes in it just in 30 minutes. It probably takes much time and effort for a non-native English speaker to develop such a competence that allows him/her to communicate accurately from the sketch.

The results of the study are also in line with Jackson (2008) that even proficient second language users might have difficulty demonstrating explicit knowledge of its particular structures during online processing. In other words, unless there is enough time to monitor and/or double-check the appropriateness of a certain structure during communication, it is difficult to be accurate. Language is mainly a means of communication, and what matters in communication is the intelligibility between interlocutors and meaning transfer, which is not always accomplished grammatically.

Richards and Renandya's (2002) idea 'that many learners assume they know the language because they know grammar and pass language tests, but in practice they discover that they lack the vital elements' supports the research findings. It is not only grammar that comes to play in communication. In practice, grammar knowledge is only a part of what is needed for an effective communication to take place. Just the same, lack of grammar knowledge may result in miscommunication because meaning, function and form are intertwined elements interacting with one another in any communicative activity.

As mentioned in chapter two, Decarrico and Larsen-Freeman (ND) have found out that many language learners can satisfactorily complete grammar

exercises when their attention is focused on the grammar, but as soon as the attention shifts to communicative interactions, the grammar is forgotten. This is completely verified by the current study. The researchers believe that examinees who answered the SWE section's items were only focused on the structure or the grammar of English, without having to worry about interaction. Therefore, they could answer most of the questions with the reference to their knowledge of language rules. In other words, the test did not require any sort of language production on the side of the examinees, thus, they were only concerned with a non-communicative practice. However, in the TWE, the focus shifted to putting language into use, i.e. using language to interact with someone through the written mode, and it was where the grammatical mistakes were observed; most of which were the mistakes that had been noticed and identified in an earlier section of the same test with the same participants.

The study suggests that almost always accuracy is better observed in grammar tests than in real communicative activities like writing compositions. Thus, the accuracy of the communicative uses of language needs to be assessed separately; i.e. in a language test consisting of vocabulary, grammar, reading, listening and writing sections, accuracy in the writing passage should have a separate share in the overall evaluation of that skill, and it should not be ignored at the price of attending to choice of words, organization of ideas, coherence, cohesion, spelling, etc. just because the grammar section of the same test specifically tests the examinee's accuracy.

Grammar knowledge has to be put into use when comprehending or producing language, therefore, it is probably best to test it in use. Many grammar knowledge tests try to isolate grammar from language use, decontextualize a sentence, and focus on a certain structural rule within a given statement. If language testing acts as a complementary practice to language teaching, the researchers believe that it is more beneficial to assess learners' knowledge of grammar by challenging their ability of using this knowledge in meaningful language production. Grammar items in a language test encouraging examinees to apply a certain structure in order to make a certain meaning can better assess examinees working knowledge of grammar or applied knowledge of language forms. Grammar items requiring examinees to recognize a grammatical mistake in a given sentence, like the SWE section in TOEFL PBT or many other similar tests, could be answered by attending to grammatical clues; this can be extensively exercised by studying sample tests resulting in examinees' getting test-wise, but not necessarily able to use language forms appropriately when producing language.

It is not advisable merely to rely on the accuracy level demonstrated in an individual's writing sample, for instance, to assess his/her grammaticality in language production stage because not everyone necessarily uses various grammatical patterns he/she knows in a single writing sample; moreover, many people use the avoidance strategy while writing or speaking; i.e. using alternative structures to convey their message instead of a certain structure they cannot use well. Therefore, in order to encourage examinees to produce a meaningful utterance, prevent their using avoidance strategy and make them use a certain structure we

intend to assess, grammar tests could be designed to ask examinees to answer certain questions using a particular structure; for example:

- (Question) Combine the following sentences into one using relative clauses:
(1) Mr. Smith has only a bother. (2) His brother is my close friend. (3) Mr. Smith lives in Riverview residential complex.
- (Answer) Mr. Smith, whose only bother is my close friend, lives in Riverview residential complex.

In the example above, the examinee needs to use a specific structure (relative clauses) to make a single sentence; therefore, he/she cannot use *coordination*, for instance, to make a compound sentence instead of *subordination* (avoidance strategy). Thus, in the example above, not only the examinees' working grammar knowledge of relative clauses could be tested, but also this very example shows whether or not the examinees know the difference between a defining and a non-defining relative clause; so even if he/she happens to combine the given sentences using relative clauses like this:

* Mr. Smith whose only brother is my close friend lives in Riverview residential complex.

We can conclude that he/she does not use defining and/or non-defining relative clauses appropriately. So, if questions like the example above are designed carefully, they can test several aspects of grammar knowledge conjointly. Now let's compare this with a classic SWE type grammar item, asking the examinee to find the mistake:

- Mr. Smith, who his brother is my close friend, lives in Riverview residential complex.

A

B C

D

Apart from the negative influence the item might have on the examinees' learning a malformed sentence, it is not that difficult to say which underlined part of the sentence is incorrect. In the researchers' opinion, the first question, which requires the use of grammar knowledge to produce a sentence, seems more challenging and beneficial for learning English.

The researchers also believe that the frequently observed mistakes in the writing samples of the learners, who had taken a comprehensive grammar course, could be used to enrich grammar teaching and writing practice in Iranian EFL environment. Some of these errors, as discussed in chapter four, are due to learners' L1 while others are not. Knowing these potentially difficult or problematic structures can come in handy when teachers teach grammar, so more elaborate examples could be set, and better explanations could be provided when teaching them; preemptive awareness might be given to students even, if needed, to reduce the risk of inaccuracies in language production.

Reference

- Brown, H. D. (2001). *Teaching by principles*. New York: Pearson Education Company.
- Brumfit, C. J. (1980). *Problems and principles in English teaching*. Pergamon Institute of English. Great Britain
- Bussmann, H. (1996). *Routledge dictionary of language and linguistics*. London: Routledge
- Crystal, D. (2004). *In word and deed. TES teachers*. Retrieved from <http://www.tes.co.uk/article.aspx?storycode=393984>
- Cele-Murcia, M. (2001). *Teaching English as a second or foreign language* (3rd ed.). Boston: HEINLE & HEINLE.
- Ellis, R. (2008). Investigating grammatical difficulty in second language learning: Implications for second language acquisition research and language testing. *International Journal of Applied Linguistics*, 18(1).
- Farhady, H. et al. (2006). *Testing language skills from theory to practice*. Tehran: SAMT.
- Hadley, A. O. (2003). *Teaching language in context*. Boston: Heinle and Heinle Publishers.
- Jackson, C. (2008). Proficiency level and the interaction of lexical and morphosyntactic information during L2 sentence processing. *Language Learning*, 58(4).
- Kathleen, S., & Kenji, K. (2010). Testing grammar. *The Internet TESL Journal*. Retrieved from <http://iteslj.org/Articles/Kitao-TestingGrammar.html>
- Purpura, J. (2010). On grammar. *Podcasts to accompany the journal Language Testing*, SAGE. Retrieved from <http://languagetesting.info/sage/ltb.php>
- Rafajlovičová, R. (2010). The status of grammar within the process of teaching and testing. Retrieved from http://www.pulib.sk/elpub2/FF/Ferencik2/pdf_doc/26.pdf
- Richards, J. C., & Renandya W. A. (2002). *Methodology in language teaching: An anthology of current practice*. New York: Cambridge University Press.
- Richards, J. C., & Schmidt, R. (2002). *Dictionary of language teaching & applied linguistics* (3rd ed.). London: Longman.
- Rimmer, W. (2006). Measuring grammatical complexity: the Gordian knot. *Language Testin*, 23, 497-519.
- Schmitt, N. (2002). *An introduction to applied linguistics*. New York: OUP.
- Siti Rohani, Z. (2007). Teaching of grammar: Teachers' beliefs, instructional context, and practices (Unpublished Ph.D. Thesis). Universiti Sains Malaysia, Malaysia.
- Widdowson, H. G. (1990). *Aspects of language teaching*. Oxford: Oxford University Press.