COMPUTER-ASSISTED ASSESSMENT:
HIGHLIGHTS AND CHALLENGES

by **Soheila Tahmasebi**

Islamic Azad University-Abadan Branch, Iran

tahmasebi_so @ yahoo.com

and **Ali Rahimi**

Bahcesehir University, Istanbul, Turkey

rahimi_cda @ yahoo.com

Abstract

Due to the unquestionable roles of technology in language classes, it might be necessary to use computers in assessing language knowledge. This study aimed to examine how computers may be used to assess language ability of ESP students. Sixty computer-major university students at Abadan University are the participants of this study. They have taken an ESP course for a four-month academic term. To measure these participants' ESP knowledge, two types of tests were used: a final achievement test based on course content, and a computer-assisted test based on TLU domain.

The study used a computer-assisted test to highlight the validity of the final achievement test. This study also investigated ESP students' perception of computer-supported assessment and highlighted some obstacles that may hinder e-supported activities in an Iranian context. Regarding the findings, the study points to the possibility of using computer-assisted assessment as an alternative to the present mainstream testing system.

Keywords: e-assessment, mainstream tests, performance-referenced tests, ESP.

1. Introduction

Attracted by many methodologists, teachers, and even students, Task-Based Language Teaching (TBLT) has not only affected teaching, but it has also positively affected assessment so much that a new paradigm of assessment has emerged and has ultimately suggested prosperous promises. Iranian EFL teachers are eagerly stimulated to devour the

merits of this rather newly proliferating change, named Task Based Language Assessment (TBLA/ TBA), endeavoring to downsize the one-shot test methods and welcoming the new trends in *alternative assessment* (Brown, 2004) to the context of English as a foreign language (EFL).

As long as the foremost purpose of including English in the instructional courses of Iranian University students is to enhance students' abilities to use language in real world situations, design, content, methodology and the evaluation system of the curriculum should support the view (Nunan, 1988). For Widdowson (2003), what students need to have acquired at the end of their EFL course is the language which is semantically encoded and pragmatically used, which consequently provides them with the ability to learn more. Therefore, in order to improve Iranian university students' abilities in using English for specific purposes – for example, English for students of computer science- the system that measures this ability should consider what will be required in the real word uses of language.

TBLA, as employed in this study, is regarded as a way of achieving a close correlation between the test performance and the criteria performance. The test performance incorporates what a testee does during the test, and the criterion refers to what the testee has to do in the real world (Bachman, 1990; Ellis, 2003).

The present study examines how TBLA measures computer-majors' abilities in particular uses of language. The participants have taken an ESP course for a four-month academic term. To estimate the validity of a typical final exam, a TBLA method is employed. After rendering a needs analysis (Munby, 1978), four tasks were selected to assess the language knowledge of the students in working with a computer. TBLA is used to seek the validity of the system-referenced test in making decisions about students' failure or success.

2. Literature review

Interestingly enough, the term 'task' has received numerous definitions and outstanding attention during its rather short history. Ellis (2003:4) gives nine definitions drawn from different research and pedagogic literature, reaching the following concept:

> A task is a work plan with a primary focus on meaning. It involves real-word processes of language use and includes any of the four language skills. Moreover, a task engages cognitive processes and has clearly defined communicative outcomes.

To have a clear overview of TBLA, the first important point seems to be introducing the notion of performance tests. Davies et al. (1999) explain that a performance test assesses participants' abilities in performing tasks, which are particularly associated to their jobs or studies. Brown (2004), on the other hand, defines task-based tests as a kind of assessment that asks for students' involvement in target language use which may not be a testing situation. The evaluation of such performances occurs according to real-world criteria. Eventually, the inferences which are made based on the results of the TBLA method cover underlying language ability or the capacity to use language use (Brindley, 1994, cited in Bachman, 2002). For Brindley (1994; cited after Bachman, 2002), language proficiency includes both knowledge and ability to use language. For example, when installing a software program it is important to accomplish the task correctly. According to McNamara (1996) this is the *strong form* of performance testing.

According to Brown (2004), performance tests may deliberately aim at eliciting some specific language performances that the testers wish to assess, hence performance assessment may not necessarily be task-based. For example, composition tasks and oral interviews are typically scored based on some linguistic characteristics of writing or speaking abilities; therefore, success or failure to complete the task is not as determining as it is in task-based assessment.

Despite this overlapping dichotomy, Brown (2004) concludes that task-based testing, which used to be a subset of performance testing, has become mixed with the overall concept of language performance testing. That is to say, in some cases participants may be required to argue for their specific ideas in an open-ended writing task in which formal elements of language as well as semantic and pragmatic issues are considered. In other words, not only accomplishing a piece of writing is valued but the linguistic elements play a role as well.

The second dichotomy related to our present discussion includes system-referenced versus performance-referenced tests. Baker (1989) defines the former category as tests that measure the knowledge of language as a system, in which language use and situations are

absent and language proficiency is central while the latter deals with providing information about using language in specific contexts. Ellis (2003) explains that these tests are construct-oriented as they are based on theories of language proficiency, in contrast, performance-oriented tests are outstandingly content-oriented and based on a "work-sample approach". In other words, they aim at assessing certain type of performance, for example, students follow some steps and solve a computer problem. Deciding to categorize a typical final exam paper, we concluded that the assessment system in Iranian Universities is almost entirely system-referenced and totally lacks the other category, performance-referenced tests. Therefore, to measure specific language knowledge, students' performance in different situations should be considered and eventually assessed.

Similarly, Brown (2004:252) asserts that "large-scale standardized tests tend to be one-shot performances that are timed, multiple-choice, decontextualized, norm-referenced, and that foster extrinsic motivation". On the other hand, performance-based assessment (or performance assessment) implies productive and observable skills such as speaking and writing in content-rich tasks. For example, a written multiple-choice test will not measure students' oral proficiency appropriately; instead, the performance of the action, or samples thereof, is required to systematically evaluate the process through direct observation by a teacher and/or self and peers. That is why the evaluation system in Iranian universities desperately asks for change and innovations in this regard are more than welcome.

Following Foster and Skehan (1999), classroom activities should move toward second language acquisition research, which invest on tasks as means of language learning. Since language use and knowledge are considered inseparable, we aimed at finding some ways to directly observe students' ability to perform specific functions and accurately assess their ESP knowledge through TBLA.

The literature of language testing indicates that TBLA could be used to assess ESP knowledge. For example, Bachman (1990) identifies two broad categories considered as language knowledge; 1) organizational knowledge which embraces grammatical knowledge as well as textual knowledge, and 2) pragmatic knowledge which includes functional and sociolinguistic knowledge. Strategic competence is also added to highlight how these sources of knowledge are realized in real world situations. Since in such a model,grammatical

knowledge is not endorsed more than pragmatic knowledge, and both validity and reliability are emphasized, measuring just organizational knowledge to make decisions about particular language knowledge would not be adequate enough.

In most ESP courses in Iran, the role of pragmatic competence is overlooked; using traditional methods, teachers assess the students' ability through tests that just include grammar and vocabulary and even translation tasks to evaluate the specific language knowledge of students. The overall result is even heartbreaking since the students' failure or success is decided upon based on this type of tests. Such a method would inevitably fail to address other sources of competence: pragmatic as well as strategic competence.

The last, but certainly not the least, important characteristic of TBLA is its positive effect on teaching since tasks coordinate assessment and teaching (McNamara, 1996). In a similar vein, Brindley (1994; cited after Bachman, 2002) calls 'task-centered assessment' an approach which has advantages over other forms of evaluation since it views language as a tool, integrates evaluation into the process of learning, and provides useful feedback as well as positive washback effects. Moreover, it is more convenient to report the results of assessment through tasks. That is, if we welcome changes in evaluation systems, teaching methodology may embrace the innovations faster and more optimistically.

To put the whole discussion in a nutshell, this paper claims that system-referenced tests are not reliable to make decision about language knowledge for particular purposes, therefore, TBLA is a more useful approach not only to provide information about students' ability, but also helping to increase their capability to learn more since the distinction between language use and knowledge becomes insignificant.

3. The study

3.1. The aims of the study

Although different researchers have employed tasks to study different aspects of language learning, TBLA as a new trend is spotted with caveats, which calls for more studies especially in the area of ESP. Since ESP draws upon the idea that the kind of language ability

which students try to learn should mirror their future use of language in real situations, this study aimed at using TBLA to measure such an ability in a valid, reliable and accurate way.

To obtain this aim, we tried to argue that the scores obtained from TBLA may display students' ability to *use* language in real world, and not just measure language knowledge. The other point addressed in this study is that most TBLA research has targeted second language acquisition (SLA) while only a small proportion has dealt with ESP contexts. The present study aims to bring performance-referenced tests based on the TBLA approach in line with a system-referenced test paradigm. In other words, it hopes to explore whether the scores obtained from a system-referenced test integrating vocabulary, grammar, and reading comprehension items represent students' target language use (TLU) domain (Bachman, 1990). To state the problem targeted by the study, the following questions were raised:

1. Do TBLA results confirm or refute students' success or failure in a system-referenced test?
2. Are the students who failed the system-referenced test capable or incapable of performing tasks employed for assessing their ESP knowledge?
3. Could TBLA replace or, at least, be integrated with, the mainstream testing methods in Iranian universities?

Regarding the first two questions the following hypotheses may emerge:

1. There is no relationship between the types of scores obtained from system-referenced tests and TBLA scores.
2. The students who have not passed the system-referenced test are not successful in accomplishing tasks presented for assessing their language abilities in real word use.

To deal with the third question of the research some problems impeding TBLA are presented at the end of section 4.

3.2. Participants

The present study used data from 50 students majoring in computer science studying at Abadan Azad University, Abadan, Iran. There were 27 males and 23 females. These students had taken an ESP course for a four-month academic term. All the students had completed Basic English and General English courses as prerequisites for taking the present course.

They have been learning English for six years at school (two hours a week) during their guidance school and high school studies. The mean age of participants was 21. Their L1 was Persian in most cases and seven of them had both Arabic and Persian as their L1; that is, they had acquired Persian and Arabic simultaneously during their childhood. Most of the students did not have jobs, and only a handful of participants had computer-related jobs and they provided support through giving information about real world uses of English.

3.3. Design and procedure

In order to check the validity of mainstream methods in assessing language knowledge, the following data collection procedures were used:

1. Questionnaires to get information about needs in order to define the TLU domain were designed since there was no well-specified course content (Bachman and Palmer 1996).

2. Oral interviews were used to elicit information from computer teachers and graduate students, or those students with computer-related jobs. These interviews highlighted the areas in which they experienced problems when they needed to use language to work with a computer.

3. Some tasks from the TLU domain were selected and eventually test tasks were isolated from the TLU tasks. Seven TLU tasks were obtained, and four of these were eventually adopted due to the limitations of the study.

4. Scales for rating tasks were developed which were tightly aligned with the types of inferences to be made (Norris et. al 2000). Since the tasks were not the same, one measurement scale could not qualify for all tasks; therefore, four rating scales were made and modified after being validated against computer teachers' views and suggestions.

5. The results of the prior final exam were gathered. Ten out of 50 students who took the ESP course failed the final exam. However, 48 accepted to participate in the study; that is, most students who failed the final exam agreed to be checked subject to the TBLA method.

6. In the end, information related to the curriculum, teachers' views toward TBLA, the context of the study and the relationship between these issues were gathered through interviews, questionnaires and personal accounts. This part of the research aimed at finding some ways to improve the evaluation system of Iranian Universities.

As it went on, through thoroughly developed procedures four tasks marked qualified in order to be used for the ultimate purpose, i.e., assessing the ESP knowledge of the participants. The selected tasks had the following characteristics:

1. Students had not practiced them in their technical courses. If they had had any technical knowledge about the tasks, it might have influenced the validity of the assessment.

2. They were not translated into Persian. Students did not have problems with computer programs already translated into Persian, therefore, to target what they needed, newly released software programs and authentic tasks downloaded from the Internet were applied.

The participants were asked to follow the instructions and then complete a task. After successfully doing each procedure, students received some points. Although after accomplishing each task the end was a measurable product, the steps to complete each task were procedural and helped the rater to evaluate the process.

Data were collected at the University of Abadan, Iran, in May and January 2008. Two different types of assessment framed the shape of this study. First of all, the results of the final test developed by the teacher were used to determine the success or failure of students in the ESP course. As a kind of achievement test, this 40-item multiple-choice test was based on the course content. In the study, this type of test is referred to as a system-referenced test since it fails to measure other types of knowledge which are required by students to use language in real world contexts. This form of assessment is typical in measuring ESP knowledge of students in Iranian universities.

Another source of data were the tasks selected based on Munby's (1978) framework for needs analysis. The tasks were administered in a university computer laboratory, where each student had access to one computer and task instructions were displayed through a visualizer. Based on the task input, the allotted time ranged from 10 to 20 minutes. The

students were asked to save the accomplished task on the desktop under a newly created file bearing their names. The data were sent to the teacher's flash memories for later checking. This event was repeated during four sessions over the period of four weeks and each time students were asked to complete one task.

To ensure the validity of task selection, two computer software specialists teaching at the same university had been consulted during the selection, implementation and evaluation of task performance. Their insightful comments helped researchers assess the completed tasks. The mean of the three scores (including one of researchers) stood for the performance of each student after one session. The mean of four performance scores was the final score of the performance-referenced test, which later was used for the further analyses.

Students themselves provided invaluable information regarding the types of problems they were facing in real uses of English in their everyday life. The students who had computer-related jobs reported the areas they were in need when using English. For example, they unanimously agreed that they have problems in understanding newly-released software installations when they are not translated into Persian.

Based on the needs analysis, students' feedback and computer teachers' suggestions, seven tasks were developed, which represented the areas that were new, not yet translated into Persian, and necessary for understanding new software programs. Eventually four tasks were selected and sequenced to represent six levels of difficulty according to combination of variables representing code complexity, cognitive complexity (Ellis 2003) and length of input (the number of steps that students should follow to complete the task – see Appendix 1). The resultant scores representing students' performances on the two types of measurement were subject to different descriptive and statistical analyses.

3.5. Results

The data analysis of the study included the following statistics.

1. One-Sample Kolmogorov-Smirnov Test was used to check if the distribution of the data was normal in order to employ further statistics. Table 1, which represents the results of One-Sample Kolmogorov-Smirnov Test, confirmed that the data of the study were distributed normally and qualified to receive further analyses.

Table 1. One-Sample Kolmogorov-Smirnov Test.

	Final	Task1	Task2	Task3
Kolmogorov-Smirnov Z	.849	.804	.672	.717
Asymp. Sig. (2-tailed)	.467	.538	.757	.682

For preliminary comparisons, a mean of the sample tasks and final scores were calculated in order to check the students' performance on each task and initially describe the results. Table 2 displays the descriptive statistics of the study. According to this table, the highest mean belonged to students' performance on Task Four, in which students were asked to install the newly-released software of Visual Studio Net. However, the highest standard deviation represents students' scores on the final exam. This means that the scores were less deviant from the mean in practical tests (Tasks 1-4). Moreover, except for Task 2, the means are positively increasing for performance tests.

Table 2. Descriptive statistics.

	N	Minimum	Maximum	Mean	Std. Deviation
Final	48	4	19	12.63	3.818
Task1	48	8	19	11.81	2.507
Task2	48	5	17	10.79	3.222
Task3	48	6	19	12.10	3.116
Task4	48	7	19	12.83	3.218
task1_4	48	7.75	17.75	11.88	2.254

2. Correlational analysis was the next step of the data analysis process, which aimed at checking the correlation between the students' performance on the final test and the TBLA tasks. Table 3 depicts the results of a close correlation between final exam scores and the TBLA scores. For example, there exists a correlation of .439 between final scores and the mean of the TBLA performance. This shows that knowledge of vocabulary and grammar

influence the overall performance of the students in doing real world tasks, even though that knowledge is not sufficient per se for using language.

Table 3. Correlations.

		final	Task1	Task2	Task3	Task4	task1- 4
final	**Pearson Correlation**	1	.290*	.305*	.359*	.352*	.439**
	Sig. (2-tailed)		.045	.035	.012	.014	.002
Task1	**Pearson Correlation**		1	.177	.357*	.307*	.574**
	Sig. (2-tailed)			.229	.013	.034	.000
Task2	**Pearson Correlation**			1	.490**	.604**	.791**
	Sig. (2-tailed)				.000	.000	.000
Task3	**Pearson Correlation**				1	.441**	.777**
	Sig. (2-tailed)					.002	.000
Task4	**Pearson Correlation**						.810**
	Sig. (2-tailed)						.000

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

3. Paired t-test analysis, comparing the means of the scores obtained from the final test and the TBLA performance, shows that no significant difference is displayed. Table 4 displays the difference between means of final scores and task-based scores.

Table 4. Paired t-test analysis.

	Mean	Std. Deviation	t	df	Sig. (2-tailed)
Final-Tasks 1-4	.73958	3.47864	1.473	47	0.147

4. Results and discussion

The results that have appeared may seemingly not be so much favorable in answering research questions; nevertheless, they have highlighted interesting facts regarding individual performances.

While doing this study, we tried to show that in order to assess the language knowledge of students in performing tasks related to their purposes, final tests, which incorporate some elements of language, do not suffice. In order to make fair judgments regarding failure or success of students, system-referenced tests should be replaced or at least integrated with performance-referenced tests.

In order to discuss the issue more systematically, the three questions of the study will be presented and answered respectively. When answering the first question of the study asked, "Do TBLA results confirm students' success or failure in a system-referenced test?", no significant difference was displayed by the paired t-test analyses. Although it was hypothesized that the students' performance on TBLA tasks does not confirm their scores on final test, the results point to the opposite direction. It should be concluded that most of the students who passed the final test could accomplish the tasks as well, and those who failed the final exam were not able to use language to do the tasks. However, the individual performance of the students presents some facts which specifically provide answers to question number two.

The second question of the study aimed at considering the performance of students on each task individually. It was hypothesized that the students who have not passed the final test, which has a system-referenced format, are not successful in accomplishing tasks presented for assessing their language abilities. The results of individual performance analyses show that some of the students who failed the final exam performed very well on different tasks, and vice versa. That is to say, in real world situations students would be required to use their knowledge of language to work on some specific tasks, and when they might not be able to do so, their language scores might be questioned.

As evidenced in Table 5, although the two types of tests are highly correlated, the individual performance of the students should not be overlooked. For example, students number 5 and 11 who obtained the highest score, 19, in the final exam could be considered, with the means of their performances on TBLA tasks being 12 and 12.25, respectively. Students number 23 and 36 performed similarly, getting 16 from the final exam and the means of 9.5 and 11.75 in the performance test. The discrepancies are not always moving in the negative direction, i.e., some students appeared capable of accomplishing the

performance tasks; hence, the case for the students who failed the course is more vital. Five out of ten who failed the course have means above ten in performing those specific tasks which required their language knowledge. For example, student number 4 failed the course; he participated in the study and got a mean of 11.75; if assessed through the TBLA method, he and four other students may have not failed the course.

Table 5. Individual performance on final exam and tasks.

N	Final	Task 1	Task 2	Task 3	Task 4	Tasks 1-4 (Mean)	Difference between tasks and final tests
1	10	8	11	10	9	9.5	0.5
2	6	10	6	7	8	7.75	-1.75
3	15	13	14	11	16	13.5	1.5
4	7	11	8	15	13	11.75	-4.75
5	19	13	8	14	13	12.0	7.0
6	14	9	10	12	15	11.5	2.5
7	15	8	15	17	18	14.5	0.5
8	12	9	11	10	12	10.5	1.5
9	15	15	8	13	15	12.75	2.25
10	13	10	10	11	13	11.0	2.0
11	19	9	15	14	11	12.25	6.75
12	18	14	16	15	19	16.0	2.0
13	18	17	15	12	16	15.0	3.0
14	14	10	13	11	10	11.0	3.0
15	14	13	9	9	12	10.75	3.25
16	15	16	11	13	12	13.0	2.0
17	13	14	6	12	8	10.0	3.0
18	11	12	9	9	14	11.0	0.0
19	13	11	15	7	17	12.5	0.5
20	6	12	5	6	15	9.5	-3.5
21	12	8	7	8	10	8.25	3.75
22	16	12	10	14	9	11.25	4.75
23	8	10	7	7	9	8.25	-0.25
24	16	12	5	11	10	9.5	6.5
25	14	12	9	9	10	10.0	4.0
26	12	10	7	12	9	9.5	2.5
27	13	9	6	12	12	9.75	3.25

28	6	11	9	7	7	8.5	-2.5
29	15	10	16	11	17	13.5	1.5
30	12	12	12	16	15	13.75	-1.75
31	17	9	11	10	13	10.75	6.25
32	18	14	14	19	18	16.25	1.75
33	14	13	13	16	17	14.75	-0.75
34	9	12	15	16	15	14.5	-5.5
35	4	14	12	12	13	12.75	-8.75
36	16	13	12	13	9	11.75	4.25
37	7	12	13	13	9	11.75	-4.75
38	6	9	15	14	15	13.25	-7.25
39	18	19	17	16	19	17.75	0.25
40	12	14	10	14	12	12.5	-0.5
41	11	12	10	14	8	11.0	0.0
42	12	11	13	15	15	13.5	-1.5
43	9	8	12	11	12	10.75	-1.75
44	15	12	9	10	13	11.0	4.0
45	10	12	8	7	12	9.75	0.25
46	13	17	13	17	15	15.5	-2.5
47	14	13	10	15	16	13.5	0.5
48	10	13	8	14	11	11.50	-1.50

While as test developers and teachers are trying to be fair and sensitive to individual differences, this important issue is worth paying much more attention to. Although final tests which try to measure language knowledge of the students are useful methods for estimating the overall understanding of the course content, they might not be valid and reliable per se for making high-stakes decisions about students' failure or success in ESP courses. The results of the study suggested that the scores on the two types of the tests are close; that is to say, students with higher scores in system-referenced tests are able to deal with computer-related tasks more profoundly. Although these results confirm the null hypothesis of the study, real-world tasks do ensure authenticity and face validity while system-referenced tests fail to do so (Ellis, 2003). System-referenced tests will determine the students' overall mastery of linguistic elements. However, there is always a certain degree of error which must be taken into account. These results might further point to testing conditions which are not always ideal; it was the first time that such practices were implemented in assessing ESP courses,

thus, some further longitudinal studies may put more light on the relation between these two types of testing.

Despite the results discussed so far, the present study also considered the feasibility of integration of TBLA with the mainstream testing methods in Iranian Universities. Looking at the problem optimistically, we should answer the question with a conditioned *yes*. To justify this suggested answer, some of the problems which are present in the Iranian teaching context in general and assessment trends in particular need to be discussed.

Different impediments are blocking the move toward innovative methods of teaching and assessing. The major obstacles to the implementation of TBLA in Iranian universities are the educational system, the teachers, the students and lack of financial support.

1. The educational system in Iran is centralized and controlled by the Ministry of Education, which determines fundamental standards for schools from kindergarten to high school in order through an educational curriculum that establishes standardized education throughout Iran. The objectives, goals, curricula, university calendar and course subjects at each different stage are formed under the laws and ordinances developed by the central government. Although teachers have the power to choose the course book that favors their interests or meets the needs of their students, the most fundamental decisions are made in advance. For example, ESP courses are regarded as theoretical; students should sit in cramped classes and digest the teacher's lectures which mostly incorporate grammar explanations and vocabulary lists.

2. The above-mentioned educational system has negatively affected students' study habits and expectations. They are accustomed to a paper-and-pencil method in which a crowded class is fronted by an active teacher who explains all the rules. Good students are supposed to listen carefully and make as many notes as possible to be accumulated for studying before the final exam (Brown, 2004).

3. The classes are overcrowded. The ideal number of students for language classes in our university is 30. However, the ESP classes are cramped with almost 70 students taking the same course and favoring one teacher.

4. It is not possible to implement the TBLA approach with all students since each student needs a computer set to carry out some real world tasks that he will be facing in the future. Thus, assigning some practical courses to ESP calls for much time and much more money.

5. Some teachers are not aware of new methods either in teaching or in testing. Even if they are, they do not feel the enthusiasm or pressure to employ what they have learnt theoretically during their own studies. Furthermore, implementing innovations requires a determined teacher as it demands a lot of time and energy on his or her part. Instead, teachers simply take the safe side and follow the mainstream methods of testing or teaching.

5. Conclusions and suggestions for further research

The required ESP courses, which are presented to the students, aim at making them ready to use language in real world situations. Vocabulary, grammar and other components of language which constitute the overall content of language tests do not measure the students' ability to use language to manipulate tasks related to their field of study and jobs. In other words, the knowledge of language should be directed toward real uses of language. Students may learn an array of vocabulary and different forms of verbs and structure and will luckily pass the final test; however, such burdens are useless as long as they do not represent the students' language ability.

Although Ss' performance on a system-referenced test and a performance-referenced test were correlated, individual performances showed some differences. In other words, having knowledge of the language system is necessary but not sufficient; it is useful only when integrated with pragmatic and strategic knowledge (Bachman, 1999:99).

Curriculum writers and materials developers should pay close attention to this issue since ESP courses are regarded as theoretical courses which should be lectured in the classroom like literature or mathematics. No need is felt to have the ESP classes in computer sites where the real world uses of language might be simulated, and real world tasks are presented to be accomplished by the students. The TBLA as an ongoing process, first, provides effective washback to modify instruction and course contents. Second, it integrates the newly released software, which is mostly in English and often problematic for students. Third, it puts forward immediate feedback to the students while they are working on tasks;

and fourth, it presents an authentic and friendly environment where students feel at ease since they have access to the teachers when they are facing some problems. These points are just a few advantages obtained after implementing TBLA.

As confirmed by the questionnaires, the immediate benefit of employing TBLA is students' satisfaction and encouragement. The second part of data collection, task implementation, was completed after the final exam, the time when students were not required to take part in the classes; thus, it was surprising that almost all students agreed to participate in the research and do the assessment tasks, showing friendly participation and invaluable help. Thus, it can be concluded that in order to measure ESP knowledge both system-referenced and performance-referenced tests should be used.

Finally, it can be stated that TBLA positively affected Ss to learn English since they voluntarily participated in this study after they had passed or failed their ESP course.

Since this study is a preliminary research attempt, directions for future explorations might be the following:

1. While this study dealt with assessing the ESP knowledge of Ss majoring in computer through some real word tasks, further studies can do the same for other fields. For example, students majoring in nursing or midwifery face similar problems when they want to use language in order to manipulate some real world tasks related to their jobs.

2. Investigating long-term effects of TBLA on more samples may lead to more assuring findings. The time and sample population were rather limited. In order to get more valid results and spot more problems, longitudinal studies are needed to deal with the effects of TBLA on evaluation systems.

3. It was almost impossible to integrate TBLA with TBLT in this study, which could be attempted in the course of further research. In order to bring about innovation in the educational system, teaching and assessment should match each other. The results of this study imply that teaching methodologies in Iran also require some movement towards the new trends in teaching.

References

Bachman , L. F. (1990). *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.

Baker, D. (1989). *Language Testing; A Critical Survey and Practical Guide*. London: Edward Arnold.

Brown, J. D. (2004) Performance assessment: existing literature and directions for research. *Second Language Studies 22*(2), 91-139. Retrieved from www.hawaii.edu/sls/uhwpesl/22(2)/Brown.pdf.

Brown, H. D. (2004). *Language Assessment: Principles and Classroom Practices*. Harlow: Longman.

Ellis, R. (2003). *Task-based Language Learning and Teaching*. Oxford: Oxford University Press.

Hudson, T. (2005). Trends in assessment scales and criterion-referenced language assessment. Annual Review of Applied Linguistics *25,* 205-227.

McNamara, T. (1996). *Measuring Second Language Performance*. London: Longman

Foster, P., and P. Skehan (1999). The influence of planning and focus of planning on task-based performance. *Language Teaching Research 18*(3), 299-323.

Munby, J. (1978). *Communicative Syllabus Design*. Cambridge: Cambridge University Press.

Norris, J.M., Brown, J.D., Hudson, T.D. and W. Bonk (2002). Examinee abilities and task difficulty in task-based second language performance assessment. *Language Testing 19* (4), 395-418.

Nunan, D. (1988). *Syllabus Design*. Oxford: Oxford University Press.

Widdowson, H.D. (2003). *Defining Issues in English Language Teaching*. Oxford: Oxford University Press.

Appendix A: Tasks 1-4

Task 1: Password protect a worksheet or workbook

Protect worksheet elements

Protect worksheet elements from all users

1. Switch to the worksheet you want to protect.

2. Unlock any cells you want users to be able to change: select each cell or range, click **Cells** on the **Format** menu, click the **Protection** tab, and then clear the **Locked** check box.

3. Hide any formulas that you don't want to be visible: select the cells with the formulas, click **Cells** on the **Format** menu, click the **Protection** tab, and then select the **Hidden** check box.

4. Unlock any graphic objects you want users to be able to change.

Task 2: Remove or change a password

Security Use strong passwords that combine upper- and lowercase letters, numbers, and symbols. Weak passwords don't mix these elements. Strong password: Y6dh!et5. Weak password: House27. Use a strong password that you can remember so that you don't have to write it down.

1. Open the document.

2. Enter the password when prompted.

3. On the **Tools** menu, click **Options**, and click the **Security** tab.

4. In the **Password to open** box or the **Password to modify** box, select the placeholder symbols (usually asterisks) that represent the existing password.

5. Do one of the following:

   o   To remove the password, press DELETE, and then click **OK**.

   o   To change the password, type the new password, and then click **OK**.

   If you changed the password, reenter the new password, and then click **OK**.

**Task 3: Wrap text around pictures in Microsoft Word – Video Tip**

If you use Microsoft Word to create any kind of document that includes pictures, wrapping text around a picture can be frustrating. Today's tip will give you some easy-to-follow steps that will have you wrapping text like a pro. A bonus of using this tip is that you can also move your pictures around your document to almost any position.

As is my cardinal rule of word processing, type all your text in first then add pictures after your text is finished. To insert a photo or other image into Word:

1. Position the cursor where you want the picture to appear

2. Click Insert –> Picture –> From File (note that you can also scan pictures directly in at this point as well)

3. Navigate to your photo

4. Click Insert

Task 4. Resize your photo to the size you need
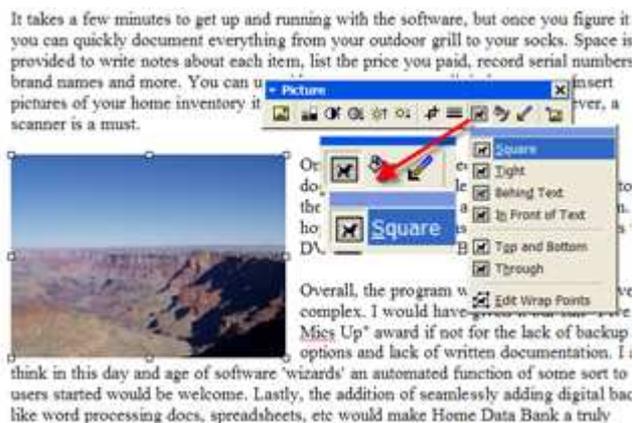
If you click ONCE on your photo, the eight 'handles' (little black squares) should appear around the picture. At this point, you should also see a small toolbar titled "Picture". If you don't see the toolbar, RIGHT click on your picture and click "Show picture toolbar". With the picture toolbar showing, follow these steps to wrap the text around the picture:

1. Click ONCE on the picture of the little dog (see illustration)

2. From the choices that appear, click the Square option

> That's it, now your text should be wrapped around the image. As I mentioned at the beginning of the tip, the bonus of following these last two steps is that you can click in the *middle* of your picture and drag it anywhere in your document.