TELL, Vol. 5, No. 1, 2011, pp. 21-43

The washback effects of ESP tests on teaching ESP at Iranaian universities

Gholamreza Kiany
Associate Professor, Tarbiat Modares University
Goodarz Albakhshi[1]
Ph.D. in TEFL, Tarbiat Modares University
Ramin Akbari
Assistant Professor, Tarbiat Modares University

Received August 4, 2009
Accepted January 3, 2010

Abstract

High-stakes tests are often used as agents for change, but attempts to introduce change in the classroom are not as effective as their designers wished they would be (Wall, 1997). The English for specific purpose (ESP) tests in Iran are such a case. The primary function of ESP tests in Iran is to select candidates for institutions of higher education (master and doctoral degrees). They are also designed specifically to promote changes in teaching ESP/ English for academic purposes at Iranian universities. This study was set out to examine washback effects of these tests on teaching ESP at Iranian universities. In doing so, data were collected through a questionnaire and classroom observations. The questionnaire was administered to 45 subject specialists teaching at Iranian universities. 10 other ESP teachers were also observed. The data were analyzed through descriptive and inferential statistics. Surprisingly enough, the results

[1] Corresponding author at: Tarbiat Modares University
Email address: alibakhshi_goodarz2000@yahoo.com

indicate that the ESP tests have fallen short of the goal. That is, these tests do not lead to innovation in teaching ESP, not do they influence teachers' teaching activities and ESP contents.

Keywords: ESP tests, washback, teaching ESP, ESP teachers

1. Introduction

Since a few decades ago, there has been a general consensus that high-stakes tests produce strong Washback (Baily, 1996; Luxia, 2005). High-stakes tests are those whose results are used to make important decisions that immediately and directly affect the test takers (Luxia, 2005; Madaus, 1990; Shohamy, 1992; 1993a, 1993b) and other stakeholders such as teachers who are helping the test takers to pass the tests, and the other participates who are engaged in curriculum development and course designing (Baily, 1996; Spolsky, 1994). It is widely accepted that high-stakes tests produce washback, in the sense that they influence the attitudes, behavior, and motivation of teachers, learners as well as parents (Pearson, 1988). Washback, an aspect of impact, has been of particular interest to both language testing researchers and practitioners and most discussions of his have focused on processes (learning and teaching). These processes occur in and are implemented by individuals, as well as educational and societal systems, and society at large. It has been discussed in language testing largely as the direct impact of testing on individuals, teaching, and learning (Hamp-Lyons, 1997a, 1997b; Hughes, 1989; Shohomy, 1993, 2001). To simply put it, language tests could have both negative and positive impacts on test takers, teachers, society and educational system.

Scholars have approached the washback effects of large-scale testing programs on education from different perspectives. Some have researched the existence and pattern of washback and the findings were mixed. However, the mixed results of the studies carried out to investigate the washback effects of tests indicate that

a group of participants are either positively or negatively influenced by language tests especially high stakes tests. Among the participants who are greatly influenced by language test results are language teachers who are the front line conduits of the washback process related to teaching (Andrews, Fullilove, & Wong, 2002). Wall and Alderson (1993) argued that one of test key characteristics is the careful observation of teacher behavior. In several of their restatements of the washback hypothesis, they emphasize the importance of teachers in washback processes. In a sense, they postulate that a test influences teaching, what teachers teach, how teachers teach, the rate and sequence of teaching, the degree and depth of teaching; and attitudes to the content, method, etc. of teaching and learning.

Among many important results of the Sri Lankan impact study, Wall and Alderson (1993) have made the following summary statements about the impact of the new Sri Lankan texts and tests on the teachers:

1. A considerable number of teachers do not understand the philosophy behind the textbook. Many have not received adequate training and do not find that the *Teacher's Guides* on their own give enough guidance.
2. Many teachers are unable, or feel unable, to implement the recommended methodology. They either lack the skills or feel factors in their teaching situation prevent them from teaching the way they understood they should.
3. Many teachers are not aware of the nature of the exam- what is really being tested.They may never have received the official exam support documents or attended training sessions that would explain the skills students need to succeed at various exam tasks.
4. All teachers seem willing to go along with the demands of the exam (if only they knew what they were).
5. Many teachers are unable, or feel unable, to prepare their students for everything that might appear on the exam.

Lam (1994) also has investigated teachers' perceptions of changes brought about by the revisions in a national exam in Hong Kong (the Revised Use of English test). He focused specifically on what he called methodology washback- that is, trying to investigate *how* teachers teach English. Building on the multiple washback hypotheses posed by Alderson and Wall (1993), Lam wanted to investigate the possibility that the revised exam would "influence how teachers teach, i.e., the methodology and methods they use to prepare students for the public examination" (1994, p. 88). He surveyed 33 teachers who had taught under the syllabuses for both the old exam and the new exam, and 28 younger teachers who had taught only under the syllabus for the new exam. Among other things, the teachers who had worked under both systems were found to be much more test-oriented than their younger counterparts (Alderson & Wall, 1993). Lam concludes that it is not sufficient to change exams: the challenge is to make a change in the teaching culture, to open teachers' eyes to the possibilities of exploiting the test to achieve positive and worthwhile educational goals (Alderson & Wall, 1993).

Andrews (1994a) has also used innovation theory as a guiding framework when he reviewed the literature on the relationship between examinations and teachers' curricular innovations. He notes that there have been instances where efforts to make language teaching more communicative have been negatively influenced by "the perceived incompatibility of such an approach with prevailing examination practices" (p. 52). One important point he makes about the potential influence of exams to bring about (or prevent) curricular and methodological change is that "examination reform may indeed be a necessary condition for educational change; it is not, however, a sufficient condition."

In the same vein Andrews (1994b) has conducted questionnaire research involving teachers in the Hong Kong context. His approach was to survey the members of the working party that revised the exam, as well as secondary school teachers affected by the change. Thirty of these teachers had taught before the introduction of the oral component in the revised exam, and had thus had experience

with both versions of the test, while 62 had not taught prior to the introduction of the oral component in the revised exam. Andrews found that both the teachers and the test developers emphasized teachers' willingness to devote time to improving students' speaking skills. However, the teachers felt that the impact of the new syllabus on the students' confidence and proficiency was not as strong as the test designers had hoped.

Boyle and Falvey (1994) have stated that there has recently been a strong link between well teaching and testing well.They also note that washback, along with validity, reliability and practicality, is now one of the main considerations in evaluating the worth of a test. Hughes (1988) has described the reactions of Turkish university English teachers to the planned implementation of a new English test that had been based on a needs analysis of the learners' intended uses of English at the university:

> The first result of even threatening to introduce a test of this kind was to cause concentration amongst the [program's] teachers. They argued that their students could not possibly cope with such a test. Pointing out that the test would actually require the students to perform just the kind of tasks that they would meet in their first year as undergraduates (and thus the kind of task for which they, the teachers, had always been preparing them) was not very much appreciated. Many teachers were convinced that they were quite unable to provide the necessary training. (p. 143)

As Shohamy, Donitsa-Schmidt, and Ferman 1996) believe, high-stakes tests are commonly practiced worldwide to change teaching and learning, especially in countries with centralized educational systems. The assumption is that due to important consequences involved, such high-stakes tests are able to make changes in teaching and learning in the ways policymakers and test designers prefer. To simply put, these tests are believed to have the ability to achieve certain positive effects on teaching and learning.

Despite the importance of washback in education and applied linguistics particularly the washback of high stake tests, only were a few international and local studies carried out to investiagte the

washback effects of high stake tests on teaching English for general and specific purposes. The preset study is an attempt to investigate washback effects of English for specific purposes (ESP) tests on the ESP teachers' teaching activities, preferences, attitudes, and perception of locally administered ESP tests.

2. Purpose of the Study

This research was designed to investigate whether or not any washback effect could be observed in teaching ESP/EAP in universities of Iran. That is, whether such tests have any significant influence on teachers' perceptions of teaching ESP, their teaching activities, their attempts to prepare such students for the tests, etc. ESP tests are administered in Iran as both achievement and selection tests. These tests are both nationally and locally administered as the entrance examinations to master and doctoral education. The objective of the study is stated in the research question: To what extent do EAP/ ESP tests influence what ESP teachers teach and how they teach in ESP classrooms at Iranian universities?

3.Method

3.1 Participants

Two groups of 55 participants took part in the study. The first group of the participants was 45 subject specialists who were teaching ESP at both master and graduate levels. These participants were teaching ESP to students of chemistry, geography, biology, and civil engineering. All the participants were either assistant or associate professors at Tarbiat Modares, Tehran, Shahed, Chamran, and Shahid Beheshti universities. All participants have been teaching English for more than 6 terms at both state and non state universities in Iran. The second group of the participants was 10 subject specialists who were observed and evaluated while teaching ESP to their students.

3.2 Data Collection Procedure

The data of the study were collected through two different ways: a questionnaire, and class observation. The questionnaire was designed to obtain preliminary data on teachers' reactions towards locally administered ESP tests, their perceptions and understanding of the examination and what they would like to do to prepare their students for the examination. The first part of the questionnaire was designed in such a way that teachers first responded to particular questions related to their language proficiency, their teaching experience, and their major. The second part of the questionnaire was designed on a five-point Likert scale which invited teachers to comment on their present teaching situation such as teachers' teaching arrangements, the choice of textbooks and the teaching methods they employ in their schools, their perception of ESP tests, and their teaching activities to prepare the students for the entrance ESP examination to master and doctoral education. The third part of the instrument consists of two open ended questions eliciting the teachers' perceptions of negative and positive washback effects of the intended tests. The construct validity of this questionnaire was estimated through running factor analysis using SPSS software version 13. Factor analysis yielded seven factors. The internal consistency for six different variables and the whole questionnaire was also calculated. The reliability indices for different variables and the whole instrument were acceptable.

3.3 Data Analysis

The data of the study were analyzed through descriptive statistics and inferential statistics. The five-point Likert scale (strongly agree, agree, neutral, disagree, and strongly disagree) was transformed into a three point scale (Agree, Neutral, and disagree).The frequency of the participants ' responses to each item as well as the sum of their responses to all the items related to each different construct of the instrument were computed. Then, to analyze the participants' responses to the items of each variable one Chi-square test was run

to find whether there is any significant difference between the participants' responses to the items of each construct. As the number of variables which this instrument was designed to measure was seven, seven separate Chi- square tests were needed to analyze the data to investigate whether there is any significant difference between the participants' attitudes towards the sum of items constituting each trait of the instrument (p value and interval confidence for each Chi-square test were .01 and .99, respectively).

4. Results

A survey was carried out in 2008 among 55 ESP teachers who have been teaching ESP for at least six terms. These teachers were from geography, biology, chemistry, Persian literature, and Civil engineering departments of the aforementioned universities. Part of the results of the study was collected through a questionnaire and part of it was obtained through classroom observation. The results of both questionnaire and classroom observation are presented, respectively.

4.1 Washback Effects on Teachers' Perception of the Tests' Purposes and Contents

Eight items of the instrument constitute the first variable known as: the teachers' perception of the purpose and content of ESP tests. The results of descriptive analysis and Chi-square test are presented in the following table.

Table 1: Teachers' perception of the content and purpose of ESP tests

Items	Agree	Neutral	Disagree
1.The purpose of ESP tests is to give feedback to ESP teachers	6(13%)	2 (5%)	37(82%)
6. The tests tasks are representative of ESP students' target needs.	6(13%)	2 (5%)	37(82%)
7. The areas of language ability which are measured by ESP tests are consistent with those	6(13%)	2(5%)	37(82%)

that are included in ESP textbooks.			
8. The characteristics of the test tasks and the tests are consistent with characteristics of teaching and learning activities.	5(11.5%)	4(8.5%)	36 (80%)
9. The purpose of the ESP/EAP tests is to contribute to the process as well as the effectiveness of teaching ESP.	6 (13%)	2 (5%)	37(82%)
10. The purpose of ESP/EAP tests is to diagnose the areas of ESP which the learners need further instruction	9(20%)	1(2%)	35 (78%)
20. The contents of ESP tests are not consistent with the contents of international EAP tests.	5 (11.5%)	2 (5%)	38 (84%)
22. Locally administered tests are needed to select the most qualified candidates for post graduate studies	5(11.5%)	2(5%)	38(84%)

(X^2=388.659, df=2, Sig=.000)

As the results shown in the above table indicate , there is a significant difference between the teachers' responses to the items measuring their perception of the test purposes and contents. That is, they disagree with all of the above eight items. In other words, approximtely 80% of the participants stated that they do not agree that that ESP tests give feedback to them, contribute to the process as well as the effectiveness of teaching ESP. Niether do they agree that these tests are represetative of the students target needs nor are they consistent with the contents of international EAP tests (see Table 1).

4.2 Washback Effects on Teachers' Teaching Methods

The five items related to this variable invited the participants to express their attiudes towards the the negative and positive impacts which thease tests might have on their teaching activities, the impacts of students' scores on their teaching methods, the stress caused by the students' scores, and their efforts to help their students to get high scores on the ESP high stake tests. The results of the descriptive analysis and Chi-square are tabulated as follows:

Table 2: Participants' attitudes towards ESP Washback effects on their teaching activities

Items	Agree	Neutral	Disagree
2. My method of teaching ESP is affected by locally administered ESP tests administered to select post graduate candidates	8(17.5%)	3(6.5%)	34 (75.5%)
5. My students' scores on achievement and high-stake ESP tests influence my teaching strategies.	9 (20%)	0 (0%)	36(80%)
16. Students' low scores on ESP tests cause me a lot of stress.	7 (15.5%)	0 (0%)	38(84.5%)
18. The locally administered ESP tests influence my teaching activities negatively.	6 (13.3%)	1(2.2%)	38 (84.5%)
33. do my best to do any activity which I find contributing to the students' scores on ESP tests.	7 (15.5%)	2(4.5%)	36 (80%)

(X^2=489.117, df=2, Sig= .000)

As the results of descriptive statistics and Chi-Square test indicate, about 80% of the participants disagree with the above items. That is, the participants do not believe that ESP tests may have positive and negative impacts on their teaching activities. Neither do they agree that they are worried about the students' achievements.

4.3 Washback Effects on Testing Preparation Activities

The items related to this category were all asked to elicit the participants' reactions to teaching activities which they do to familiralze their students' with content of the tests, test task types, as well as the strategies which the test takers may need to answer test items more easily and correctly. The results (descriptive and inferential) are presented in the following table**.**

Table 3: Participants' responses to items related to testing preparation activities

Items	Agree	Neutral	Disagree
4. I familiarize my students with the content of ESP tests.	6(13.5%)	0	39(86.5%)
11. I always review the yearly local ESP tests so that I can make a change in my teaching activities in the classroom settings.	7 (15.5%)	0	38 (84.5 %)
24. I teach my students the techniques needed to answer multiple choice questions.	9 (20%)	2 (4.5%)	34(75.5 %)
32. I always distribute sample test items among my students.	8 (18%)	0	37(82%)

(X^2=489.117, df=2, Sig= .000)

As the results in the above table indicate about 86.5% of the participant disagreed that they familiarized the students with the content of ESP tests. 84.5% disagreed that they always reviewed the yearly local ESP tests so that they can make a change in their teaching activities in the classroom settings.82% disgreed with the themes of distributing sample tests among the students and teaching the techniques needed to do the tests tasks to their students. Moreover, the results obtained from classroom observation indicate that none of the 10 participants observed attemted teaching to test activities such as distributing sample questions, etc.

4.4 Teachers' Attitudes towards ESP Tests' Contents

The participants were also asked to express their attitudes towards the main contents of the tests and the domains which they must in fact measure. The descriptive analysis of the data followed by a Chi-square test is presented in the following table.

Table 4: Teachers' reactions to ESP tests

Items	Agree	Neutral	Disagree
12. In addition to reading skills, the other language skills should be evaluated	36 (80%)	4 (9%)	5(11%)
13. ESP tests should only aim at helping the learners to use language to meet their own academic needs.	34 (75%)	4 (9%)	7(16%)
14. Instead of being used to select master and doctoral students, ESP tests should be used to place the students into the right channel of language learning.	35(78%)	5(11%)	5(11%)
17. ESP tests used in our local universities are not reliable enough to rely on.	31(69%)	8(17.5)	6(13.5)

(X^2=125. 568, df=2, Sig= .000)

As the results indicate, the participants agrreed with the above items. That is, about 75% believe that the contents of the ESP tests do not cprrespond with the academic skills needed by the master and doctoral students. About 12% were neuteral and 13% disgreed with the contents of the above items. The results of Chi- square, as shown in the above table, indicate that there is a significant difference between the frquency of responses given to the items measuring the above variable. That is, the frequency of "agree" is significantly greater than the other two responses.

4.5 Washback Effects on Teaching Contents

Washback influence on teaching contents has been more frequently documented by the empirical studies. Seven items are asked to investigate ESP tests washback influence on ESP teaching contents. The results are presented in the following table.

Table 5: The results of washback effects on teaching contents

Items	Agree	Neutral	Disagree
19. Although the tests do not measure listening, writing and speaking skills, I do my best to	6(13.5)	2(4.5)	37(82 %)

allocate time to teach of them.			
25. In addition to their ESP textbooks, I motivate my students to read the journals which they find related to their fields of study.	7(15.5%	2(4.5%)	38 (80%)
26. I will provide my students with educational software as supplementary materials.	9 (11.5%)	0	36(79.5%)
27. In addition to university ESP classes, I encourage my students to attend the other language classes.	8 (18%	0	37(82%)
29. In addition to language general skills, I teach the other academic skills such as note taking and summary writing.	7 (15.5%	0	38 (84.5%)
30. I remind my students that they need academic skills such as writing ISI papers and listening to lectures to attend international conferences.	8 (18%	3 (6.5%)	37(75.5%)
31. In addition to general knowledge skills, I use original specific texts	8(18%)	3 (6.5%)	37(75.5%)

(X^2=317.505, df=2, Sig= .000)

It can be seen from the table 5 that approximately more than 70% of the participants disagrred with the items measuring washback effect of ESP tests on teaching contents. The results of Chi-square tests also indicates that there is significant difference between the participants' responses to the above items(p=.01). That is, they believe that they do not teach language skills except technical reading, niether do they provide their learners with the the other hardware and software supplementary materials which their students may need to be proficient in.

4.6 Teachers Perception of Their Familiarity with ESP Tests

Two items were designed to elicit the participants' ideas about their familiarity with local ESP tests contents and the skills which are measured by these tests. The results are represented in the following Table.

Table 6: Teachers perception of their familiarity with ESP tests

Items	Agree	Neutral	Disagree
3. I am familiar with the local ESP tests contents.	33(73%)	5(11.5%)	7 (15.5 %)
21. I know what skills are evaluated by the ESP tests.	37 (82.5%)	2 (4%)	6 (13.5%)

(X^2=80.600, df=2, Sig= .000)

As the results indicate 73% of the participants indicated that they are familiar with local ESP tests contents and 82.5% stated that they know what skills are measured. The chi-square test also indicate that there is a significant difference between the participants responses.

4.7 Teachers' Perseption of Their Abilities to Construct ESP Tests

The last component of this instrument investiagates the participants' views of their abilities to construct ESP tests and prepare the students to do the test tasks which might appear on the exam. The results are presented in the following table.

Table 7: Teachers' attitudes toward their abilities to construct ESP tests

Items	agree	Neutral	disagree
15. I am able to prepare my students for everything that might appear on the exam	5 (11.2%)	1 (2.3%)	39 (86.5%)
34. I am familiar with principles and procedures of ESP test development.	0	7 (16%)	38(84%)

(X^2=110.600, df=2, Sig= .000)

5. Discussion

The influence of examinations on teaching and learning has been discussed extensively in the general education and the language education literature (Alderson & Wall, 1996; Chapman & Snyder, 2000; Davies, 1988, 2001; Frederiksen & Collins, 1989; Madaus, 1988, 1990; Oxenham, 1984; Wall, 1997, 2000; Wall & Alderson, 1993; Wong, 2001). Since many years ago, influence of language tests was assumed to be negative, and indeed such negative perceptions are still of much concern, as illustrated by Chapman and Snyder (2000) "teachers' tendencies to teach to the test are often cited as an impediment to introducing new instructional practices"(p. 460). More recently, however, attention has switched to the possibility of exploiting the power of high-stakes tests to positive ends in support of innovation and changes in curriculum, teachers' methods, and stake holders' perception of high stake examinations. The belief underlying this strategy is summarized by Biggs (1996) as follows: the quickest and easiest way to change student learning is to change the assessment system. Despite the assumption that tests are sometimes harmful, the use of assessment as a means of promoting curriculum change has become increasingly common, in education generally (see, for example, Chapman & Snyder, 2000; James, 2000), and also in language education (see, for instance, Cheng, 1997, 1998; Pearson, 1988; Wall & Alderson, 1993).

The main question of this study was whether the ESP tests administered as entrance examination to master and doctoral education have had any washback effects on the teaching of English for specific/academic purposes at Iranian universities. The results of factor analysis indicate that the initial Eigen values for each component was above 53 and the loading of each factor was above .65 (The details of factor analysis for each skill are presented in the Appendix). A brief look at the loadings shows that almost all of the loadings are high enough to conclude that all ten items of each component constitute one factor.The results of the study are discussed under seven different themes of washback effects of the ESP tests on teachers' teaching methods, teaching contents,

perceptions of the purpose and contents of tests, test preparation activities, perception of their abilities to prepare the students for high stake ESP tests, attitudes to tests contents, and their familiarity with test construction.

In terms of the first variable of the study, washback effects on teaching methods, the results of the present study (see table 2) indicate that more than 75% of the participants stated that current ESP tests administered to select master and doctoral candidates did not have any positive or negative impacts on their teaching methods. The results of the classroom observation also indicate that almost all ESP teachers teach in the same way. The most frequent teaching activities observed were reading aloud and translation from English into Persian which were never used in ESP part of master entrance examination. Therefore, it could be strongly argued that ESP tests do not have enough power to promote teachers' teaching methods. They do not cause any stress and anxiety among the teachers and students' achievements on these high stake tests are not of much concern to the ESP teachers at all.

The second area of washback which was investigated in the present study was teaching content. According to Wall (1997) teaching content is an area of washback intensity. The results of the study (See Table Five) indicate that the participants involved in teaching ESP to Iranian master and undergraduate students do not attempt to spend time teaching the academic skills such as listening, writing, and speaking to the students. They do not motivate the students to study the other language skills which the students need in their present use situation and target language use situations. Almost 80% stated that they do not provide their students with any journals, original texts, and the other hardware and software instructional materials, either. Therefore, it could be discussed that high stake ESP tests used at Iranian universities failed to produce fundamental changes in teaching ESP contents. The results gathered through classroom observation point out that the only skill emphasized in ESP classes at Iranian universities is technical reading. All the texts used in the classroom were either selected from the books specific to the students' own field of the study such

as chemistry, geography; biology, etc or were adapted from Latin internet sites . The students' learning activities were translating the passages from English to Persian and sometimes reading the texts aloud.

In terms of ESP tests failure to produce great change in teaching contents, it could be argued that there may be different influential factors but the main reason may be lack of interactional and situational (Bachman & Palmer, 1996; Douglas, 2000) authenticity in ESP tests administered locally and internationally in Iran (Alibakhshi, Kiany, & Akbari, 2010).

Perception of the stake holders of the tests tasks and task specifications of the test is the next area of washback. The results of the study, descriptive and Chi-square test (Table One) show that almost more than 75% of the participants of the study stated that in addition to technical reading the other language skills should be evaluated. They also believe that a change in the purpose of the ESP tests is needed. That is, they should be administered as placement tests so that the language areas in which the learners need further instruction are diagnosed right before they start their master or doctoral studies. In other words, their performance on these scores should not influence their entry to universities as master/ doctoral students. Although the participants believe that ESP test items are not representative of the learners' academic needs, surprisingly enough they do not attempt to teach the academic skills needed by the learners. One possible reason, as the results obtained through classroom observations indicate, is that ESP teachers' perception of ESP is not valid. Neither do they know what ESP is, nor do they know enough about learning and teaching theories underlying ESP.

Washback influence on the teachers' perception of the test purposes and characteristics was another aim of this study. The results of Table 4(descriptive and Chi-square) point out that about 80% of the participants stated that the present ESP tests are not representative of the students academic needs. They also believe that they do not contribute to the process of teaching; neither do they give the teachers feedbacks on their own teaching effectiveness. Moreover, they stated that such tests are not needed to

select the most qualified master and doctoral students. In line with those who believe that there should be a correspondence between tests tasks and target language use situation tasks, it could be discussed that because of lack of correspondence between tests tasks and learners' academic needs, the ESP tests cannot produce any significant washback influence on teaching process.

Test preparation activities in the classroom have also been assumed to be related to washback effects on teaching process. The empirical studies indicate that the introduction of some language tests caused the teachers to teach to the test. That is, the teachers attempted to familiarize their students with different sample tests and the techniques they need to answer the test tasks easily. The results of the present studies indicate that more than 80% of the participants argued that they do not teach to ESP tests. In other words, neither do they familiarize their students with the ESP tests tasks, nor do they allocate time practicing sample ESP tests.

The last component of this washback study was the teachers' perception of their abilities to construct ESP tests and to prepare their students to answer any test tasks which they may encounter in tests of ESP/EAP. It is assumed that all teachers must be able to construct any kinds of test tasks and at least prepare the students to be capable of answering any kind of test tasks. However, the results of this study (Table 7) indicate that more that 80% of the participants stated that neither are they familiar with testing fundamental principles nor are they able able to prepare their students to do any tasks which might appear on the the ESP tests.

6. Conclusion

Viewing the results of this study, we conclude that ESP tests administered in Iran do not lead to any innovation in teaching ESP at Iranian university. As the review of the washback studies demonstrate it is simplistic to believe that a test can result in all desired changes in teaching and learning. Teaching ESP is a complex phenomenon and there are many factors involved in bringing about changes, like the school environment, messages from

administration, expectations of teachers and students. As Saif (2006) argues, an analysis of the needs and objectives of learners and educational systems should be carried out as a starting point for the research in washback. Wesche (1983, cited in Bailey, 1996) points out that when tests reflect the situations, content and purpose where learners will use the language, they are likely to improve motivation. Shohamy et al. (1996) consider factors like the status of the subject-matter tested, the nature of the test, and the use to which the test scores are put (p. 300). A list of factors which might have prevented the ESP tests in Iran from providing an effective 'lever for change' can be stated can be categorized under different themes as:

1. Teachers' factors including misunderstanding the nature of ESP, unfamiliarity with testing fundamental principles, inability to prepare their students to answer any kinds of test tasks, unfamiliarity with new approaches to teaching language, and unfamiliarity the students' language needs
2. The nature of ESP tests including contents, purposes, and authenticity
3. Learners' factors such as their language needs, motivation, and learning strategies
4. Education system factors including its attitudes to ESP as a field, curriculum development, etc.

In order to accelerate the washback influence of ESP tests, the above mentioned factors should be taken into consideration within any educational setting.

References

Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics, 14*(2), 115-129.

Alibakhshi, G., Kiany, G.R., & Akbari, R. (2010). Authenticity in ESP/EAP selection tests administered at Iranian universities. *Journal of Asian ESP*, *6*(2), 65-93.

Andrews, S. (1994a). The washback effect of examinations: Its impact upon curriculum innovation in English language teaching. *Curriculum Forum 4*(1), 44-58.

Andrews, S. (1994b). Washback or washout? The relationship between examination reform and curriculum innovation. In D. Nunan, R. Berry, & V. Berry, (Eds.), *Bringing about change in language education." Proceedings of the International Language in Education Conference 1994* (pp. 67-81). Hong Kong: University of Hong Kong.

Andrews, S., Fullilove, J., & Wong, Y. (2002). Targeting washback: A case study. *System, 30*, 207-223.

Bachman, L., & Palmer, A. (1996). *Language testing in practice*. Oxford: Oxford University Press.

Bailey, K. M. (1996). Working for washback: A review of the washback concept in language testing. *Language Testing 13*(3), 257-279.

Bailey, K. M. (1999). *Washback in language testing. Educational testing service*. Peicton: Newjersey RM-99-4.

Biggs, J. B. (1996). *Testing: to educate or to select?* Hong Kong: Hong Kong Educational Publishing.

Boyle, J., & Falvey, P. (Eds.). (1994). *English language testing in Hong Kong*. Hong Kong: Chinese University Press.

Chapman, D.W., & Snyder, C.W. (2000). Can high stakes national testing improves instruction: reexamining conventional wisdom. *International Journal of Educational Development, 20*, 457-474.

Cheng, L. (1997). How does washback influence teaching? Implications for Hong Kong. *Language and Education, 11*(2), 38-54.

Cheng, L. (1998). Impact of a public English examination change on students' perceptions and attitudes toward their English learning. *Studies in Educational Evaluation 24*(3), 279-301.

Davies, A. (1988). Communicative language testing. In A. Hughes (Ed.), *Testing English for university study: ELT Documents 127*. Oxford: Modern English Publications.

Davies, A. (2001). The logic of testing languages for specific

purposes. *Language Testing, 18*(2), 133-147.

Douglas, D. (2000). *Assessing language for specific purposes*. Cambridge: Cambridge University Press.

Fredericksen, R., & Collins, A. (1989). A system approach to educational testing. *Educational Researcher, 18*, 27-32.

Hamp-Lyons, L. (1996). Ethical test preparation: The case of TOEFL. Paper presented at the Language Testing Research Colloquium, Chicago, IL.

Hamp-Lyons, L. (1997a). Ethics in language testing. In C. Clapham (Ed.), *The Encyclopaedia of Language and Education.* (Vol. 7 Language Testing and Assessment., pp. 323-333.). Kluwer: Dordrecht.

Hamp-Lyons, L. (1997b). Washback, impact and validity: ethical concerns. *Language Testing, 14*(3), 295-303.

Hughes, A. (1988). Introducing a needs-based test of English language proficiency into an English-medium university in Turkey. In A. Hughes (Ed.), *Testing English for university study, ELT Documents 127* (pp. 134-146). Oxford: Modern English Publications.

James, M. (2000). Measured lives: the rise of assessment as the engine of change in English schools. *The Curriculum Journal 11*(3), 343-364.

Lam, H. P. (1994). Methodology washback: An insider's view. In D. Nunan, R. Berry, & V. Berry (Eds.), *Bringing about change in language education: Proceedings of the International Language in Education Conference 1994* (pp. 83-102). Hong Kong: University of Hong Kong.

Luxia, Q. (2005). Stakeholders' conflicting aims undermine the washback function of a high-stakes test. *Language Testing*, 2005; 22; 142

Madaus, G. (1990). Testing as social technology. Paper presented at the inaugural annual Boisi Lecture in Education and Public Policy. Boston College.

Oxenham, J. (1984). *Education versus qualifications?* London: Allen and Unwin.

Pearson, I. (1988). Tests as levers of change (or 'putting first things first'). In D. Chamberlain & R.Baumgartner (Eds.), *ESP in the classroom: Practice and evaluation* (pp. 98-107). ELT Documents 128. London: Modem English Publications in association with the British Council.

Saif, S. (2006). Aiming for positive washback: A case study of international teaching assistants. *Language Testing, 23*(1), 1-34.

Shohamy, E. (1992). Beyond proficiency testing: A diagnostic feedback testing model for assessing foreign language learning. Modern *Language Journal*, 76(4), 513-521.

Shohamy, E. (1993a). *The power of tests: The impact of language tests on teaching and learning*. NFLC Occasional Paper. Washington, DC: National Foreign Language Center.

Shohamy, E. (1993b). A collaborative/diagnostic feedback model for testing foreign languages. In D. Douglas & C. Chapelle (Eds.), *A new decade of language testing research* (pp. 185-202). Alexandria, VA: TESOL Publications.

Shohamy, E. (2001). *The power of tests.* London: Longman.

Shohamy, E., Donitsa-Schmidt, S., & Ferman, I. (1996). Test impact revisited: Washback effect overtime. *Language Testing, 13*(3), 298-317.

Spolsky, B. (1994). The examination-classroom backwash cycle: some historical cases. In R. Berry (Ed.), *Bringing about change in language education: proceedings of the International Language in Education Conference 1994* (pp. 55-66). Hong Kong: Department of Curriculum Studies, The University of Hong Kong.

Wall, D. (1997). Impact and washback in language testing. In C. Clapham & D. Corson (Eds.), *Encyclopedia of language and education, Volume 7: Language testing and assessment* (pp. 291-302). Dordrecht: Kluwer.

Wall, D. (2000). The impact of high-stakes testing on teaching and learning: can this be predicted or controlled? *System 28*, 499-509.

Wall, D., & Alderson, J.C. (1993). Examining washback: The Sri Lankan impact study. *Language Testing, 10*(1), 41-69.

Wong, Y. (2001). The impact of a new oral exam on students' spoken language performance. Unpublished Ph.D. dissertation, University of Hong Kong.