

On the Consequential Validity of ESP Tests: A Qualitative Study in Iran

*Gholam Reza Kiani
Tarbiat Modares University
Goodarz Alibakhshi*
Yasouj University
Ramin Akbari
Tarbiat Modares University*

Consequential validity, a component of construct validity as a unified concept introduced by Messick(1989), deals with the impacts of tests on teaching, learning, individual test takers, teachers, society, and educational system within a country. Although the impacts of language tests on teaching and learning have been somehow studied, the consequences of ESP tests on individual test takers and teachers have not been given appropriate attention. The present study aimed at exploring the possible consequences of ESP tests, as parts of master and doctoral entrance examinations to post graduate universities in Iran, on the stake holders and society. In doing so, a qualitative research method was applied. The participants were 31 master and doctoral students, and 5 ESP instructors. The data were analyzed in terms of the content. Four main themes with a variety of sub-themes emerged from the content analysis. The results indicated that ESP tests have negative psychological (anxiety, stress, disappointment, self-confidence, teaching efficacy, disappointment), social (deprivation from education, ethical issues, acceptance of nonqualified candidates), financial, and family consequences on the learners and teachers. The society is also influenced as the result of ESP tests scores.

*Corresponding Author, Email Add.:

alibakhshi_goodarz2000@yahoo.com

Keywords: Validity, Consequential Validity, ESP Tests, Language Teaching

The consequential aspect of construct validity has been defined in different ways. According to Messick(1989) consequential validity includes evidence and rationales for evaluating the intended and unintended consequences of score interpretation and use in both short- and long term, particularly those associated with bias in scoring and interpretation, with unfairness in test use, and with positive or negative washback effects on teaching and learning. However, this form of evidence should not be viewed as a separate type of validity, say, of 'consequential validity' or, still worse, 'washback validity' (Messick, 1998). Bachman (1990) and Bachman & Palmer (1996) used the term *impact* to describe these consequences of tests. The impact of test use operates at two levels: a micro level, in terms of the individuals who are influenced by a particular test use, and a macro level, in terms of the educational system or society (Bachman & Palmer, 1996). In this study, following the taxonomy of Bachman and Palmer, the consequences of tests on teaching and learning are viewed as washback effects, and the consequences on individual stake holders, i.e., learners, teachers, parents, and test takers' family members, and society are considered as impact.

Since a few decades ago, the impacts of different language tests, either at macro or micro levels, have been studied. The literature review indicates that there has been a general consensus that high-stakes tests produce strong washback. High-stakes tests are those whose results are used to make important decisions which immediately and directly affect the test takers (Luxia, 2005; Madaus, 1988; Shohamy, 1993a, 1993b, 2001) , other stakeholders such as teachers who are helping the test takers to pass the tests, and the participates who are engaged in curriculum development and course design (Baily, 1999 ; Spolsky, 1997). The term 'backwash' has been used to refer to the way a test influences teaching materials and classroom management (Hughes,1989), although in applied linguistics and language testing community the term 'washback' is more widely used today (Weir 1990; Alderson

&Wall 1993; Alderson, 2004). Washback is generally known as being either negative or positive (Taylor ,2005) . Negative washback is said to occur when the content or format of a test is based on a narrow definition of language ability, which, thus, constrains the teaching/learning context. Positive washback is said to result when a testing procedure encourages ‘good’ teaching practice, for example, an oral proficiency test is introduced in the expectation that it will promote teaching speaking skills (Taylor, 2005).

The impact of a test can be immediate or delayed (Andrews, 1994; Andrews & Fullilove 1994). According to these researchers, washback seems to be associated primarily with ‘high-stakes’ tests, that is, tests used for making important decisions that affect different sectors, for example, determining who receives admission into further education or employment opportunities (Chapman & Snyder, 2000). Shohamy, Donitza-Schmitdt and Ferman (1996) believe that the situations in which admission, promotion, placement or graduation is dependent on the test are very important, and require great care.

According to Taylor (2005), language tests can have consequences beyond just the classroom. That is, tests and test results have a significant impact on the career or life chances of individual test takers (educational/employment opportunities). They also influence educational systems and society at large, for example, test results are used to make decisions about school curriculum planning, immigration policy, or professional registration for doctors. The growth of a test may lead publishers and institutions to produce test preparation materials, and run test preparation courses. Bachman (1990) used the term *impact* to describe these consequences of tests. Some language testers consider washback as one dimension of impact, describing effects on the educational context (Hamp-Lyons 1997); others see washback and impact as separate concepts relating respectively to ‘micro’ and ‘macro’ effects within society (Bachman & Palmer, 1996). Most testers locate both concepts within the theoretical notion of ‘consequential validity’, in which the social consequences of testing are part of a broader, unified concept of

test validity (Messick, 1989, 1996). Consequential validity has been extensively discussed among language testers in the recent years (Kunnan, 2000). As consequential validity is a complex notion, we try to study some specific parts of it known as tests consequences on test takers, teachers, and society.

Although washback effect on teaching and learning has been both theoretically and empirically discussed, and several washback hypotheses were put forward, quite a few empirical studies have been carried out to explore the ESP tests negative or positive consequences on test takers' life and society at large. Therefore, the present study aims at exploring the ESP tests' consequences on Iranian stake holders' life, and society. In doing so, the following research question was raised:

What impacts do ESP/ EAP tests have on the life of candidates for post graduate studies as well as the system of higher education at Iranian universities?

Method and Procedure

This study is part of an extensive investigation to explore the consequences of ESP tests, which are parts of master and doctoral entrance examination of state universities, consequences on master, and PhD candidates, and ESP teachers. To simply put it, the purpose of this study was to identify and describe the consequences which these tests may have on Iranian stake holders including teachers, test takers, parents and learners' family members, and society.

A qualitative approach was selected because this aspect of consequential validity is underpinned by personal and organizational culture, and little of relevance has been reported in the literature. Qualitative research is context based, thus it is imperative for researchers to recruit participants in a transparent manner. The participants were 16 master and 10 PhD students of different majors, and 5 ESP teachers at different universities in Iran.

The data were gathered during face-to-face in-depth interviews. The researchers informed the participants of the

purpose of the research, and obtained their written consent. The researchers also obtained the participants' permission to audiotape each interview for the purposes of content analysis and audit trail. The interviews were conducted in both an unstructured and a semi-structured manner. All participants were interviewed in privacy. Each interview began with the question: 'What do you think of ESP tests consequences?' the answer to which was followed by the questions designed to elicit specific consequences of the target tests, such as: 'What was the most positive or negative consequence of these tests?'

The participants were also asked to describe their experiences, attitudes and beliefs about the ESP tests, and their impacts on their life. The interviews lasted, on average, for about 30 minutes. Interviewing took place during all days over a six-week period, until the data collected were being consistently duplicated. No new information was gained from the last three interviews, thus, data saturation was considered to have been achieved.

The interview data were immediately transcribed verbatim, and analyzed using qualitative content analysis. Content analysis is a subjective interpretation of the content of textual data using a process of systematic classification. This process uses mainly inductive reasoning, by which themes and categories emerge from raw data under careful examination and constant comparison (Strauss & Corbin, 1990). One characteristic of qualitative content analysis is that the method, to a great extent, focuses on the subject and the context, and emphasizes differences between and similarities within themes and categories. Another characteristic is that this method deals with manifest as well as latent content in a text. Manifest content consists of respondents' actual words forming concepts, while themes are seen as expressions of the latent content. In this study, the method of coding according to qualitative content analysis was used to derive categories and themes from the data, which were identified from the first interviews, and then tested and revised through the analysis of succeeding interview (Marvast, 2004).

To ensure reliability and provide inter-rater reliability, the

researchers coded interviews individually, discussed the outcome, agreed on changes, and then, separately coded the next interview. In the first five interviews, over 80% of the codes were shown to be consistent between the two researchers. These interviews were re-coded after a two-day interval by the same team, and found to be stable. The same coding scheme was then applied to a re-analysis of all interviews. The researchers also reviewed and discussed the entire interview coding to ensure consistency.

Findings

The 31 participants consisted of 16 master and 10 PhD students, and 5 ESP teachers. Four themes were extracted from the interview data using qualitative content analysis: psychological, social, financial, and family consequences. Each of these main categories is further divided into subcategories which are described in details in the following parts of the study.

Psychological Consequences

Almost all participants stated that ESP tests had great psychological consequences. This main category is subcategorized into stress, anxiety, self-confidence, depression, teaching efficacy, and disappointment.

- Stress and Anxiety

Many of the participants acknowledged that they had experienced a kind of anxiety before or even after ESP tests are administered. The following examples illustrate this theme.

ESP tests are really difficult and the passages are long. When I do not know the meaning of unknown words, I become nervous and I do not attempt all the items. My performance on these tests influences my future so thinking about the result causes me stress. In fact, all tests are stressful but language tests are something different because a specific preparation is needed (participant 1).

Another participant described test consequences as stressful. He said:

The results are not announced soon. Sometimes, I have to wait for about three months. During this time, I always think about the test results. Such thoughts cause me a lot of stress (participant 3).

The PhD candidates also acknowledged that the ESP section of the entrance examination was terribly stressful. A participant said:

I had no problems in content subjects. My only problem was specific English. I almost translated about 70 pages of my ESP book but the passages given to us to be translated were unseen. When I did not know the meanings of some unknown words, I got confused. Such confusion led to my mental stress. That is why; I could not have a good performance. Even one point can make a change in the chance of admission. When the results were announced, I noticed that if my score on ESP section were one point higher, I would pass the entrance exam (participant 12).

- Depression

The majority of the participants argued that the results of the tests and test phobia greatly influenced and depressed them so that they could not study anymore. Such a kind of depression sometimes affected their daily activities. The following examples illustrate this theme.

You may not know about test consequences. The English tests sometimes turn out to be a kind of disaster in our life. Due to many known and unknown factors, we are depressed before and after taking language tests. Sometimes we feel so depressed that we cannot do anything. We cannot even get out of bed to eat breakfast or lunch. Even, we do not talk to anyone for a couple of days (participant 6).

Another participant stated:

Regardless of test results, whether to fail or pass, due to the nature of ESP tests administered at state and non-state

universities in my country, I become depressed. The tests are not measuring what I need. They only test translation ability of the students whereas I do not need translation. I may pass the test but I cannot use ESP language to meet my academic needs. My score is not bad but I cannot even write a paragraph or summarize a passage. That is why I always feel depressed (participant 16).

- Disappointment

The participants of the study believed that they sometimes felt disappointed and gave up studying. The following examples are given to illustrate this theme:

I took PhD test several times but I always failed because of my bad performance in ESP section of the test. I got tired and finally disappointed. Even if I pass the ESP part of the test, I am not pleased because these tests do not measure what we need at all. Therefore, I always felt disappointed, and planned to give up my studying for PhD examination (participant 8).

Another participant added:

At first I was really interested in studying for PhD entrance examination but having failed the test several times I lost my motivation, and felt disappointed. Such a kind of feeling lasted for a long time. Despite the unreliability and invalidity of these tests, they had great influence on the candidates' admission. I got disappointed, and unwillingly gave up PhD(participant 2).

As the participants do not know anything about the scope of the contents which are given in the test, they may not attempt hard to get prepared. They believe that their knowledge in English is limited and they may not be able to do their best in the unseen contents. One of the participants said:

The candidates do not know how much preparation is needed. The ESP texts are infinite. I remember I studied and translated more than 200 pages but the test passages were not in the range of materials which I was familiar with. Even an unknown word confused me. That is why I am not sure I can answer all the passages because they may be unfamiliar to me,

and this makes me disappointed. I gave up studying ESP (participant 10).

Furthermore, the participants acknowledged that their performance on ESP tests depends on their general language performance to a great extent. As they are not good at general language, they have no chance to learn ESP. They believed that they almost know some technical words which they cannot use in context because of their general language. As the results, they are hopeless to learn English for ESP/EAP purposes. One of the participants stated:

English language is not given appropriate attention in our country. The time allocated to teaching English at secondary and tertiary schools is not enough. Teachers only focus on grammar and reading. Because we have no exposure to authentic language, we easily forget what we learn in high schools. How does one expect us to learn English in such a kind of educational system in which only a few grammatical structures and a couple of words are taught? I studied English for about seven years but I cannot even write a short paragraph. I think the results of ESP tests indirectly disappoint the students. I myself am one of those disappointed ones. I think I never pass ESP/ EAP tests such as IELTS or IBT TOEFL. In one of the PhD tests, taken two years ago, the test takers were asked to translate a passage from Persian to English. Although I knew all words, I could not do the task because I did not know how to combine the words to make a text (participant 18).

- Self-Confidence and Efficacy

The participants believed that their self-confidence is greatly influenced by these tests. They stated that by using their short memory capacity, they can memorize a list of technical words and some specific passages. They may attain a good score which leads to a kind of false self-confidence. They may think that this score indicates their true ability, and may give up studying English. One of the participants said:

I just studied for two months. In fact, I got a good score. Having seen my score on ESP test, I felt a high self-confident. I thought I could meet my all academic needs; therefore, I stopped reading my English books. During my master classes, I noticed how weak I was in English. I could not even translate a paper. I had to allocate almost all my time studying English. You cannot imagine how difficult it was for me to pass ESP courses in two semesters (participant 17).

Concerning the impact of the tests on self-confidence, another participant stated:

When I saw my score I became very confident, and felt proud. Some of my friends failed the test. I sneered at them. It was a great achievement for me. The majority of the test takers failed these ESP tests, but I passed. Therefore, it was a good courage for me to take even difficult tests. English tests differ from other tests, so passing them successfully is really of much significance in my confidence (participant 13).

The results of the study also indicate that the students' low scores on ESP tests led to self-confidence wakening. One of the participants said:

I have taken PhD entrance examination three times. My scores on subject tests were not bad, but I had no good performance in the ESP sub-test of the entrance examination. My friends who were good at English could pass the entrance examination, and enter university. Such a failure in ESP tests caused me to lose my self-confidence, and gave up preparing for PhD examinations because I knew it does not worth the price to spent time learning English (participant 6).

The results also indicate that ESP tests will cause a kind of false positive sense of teaching efficacy among ESP instructors. The following example illustrated this theme.

The so-called ESP tests administered at our universities just lay emphasis on technical vocabulary and technical reading passages. I am good at these skills. I am able to teach these skills very well. Because of such preparedness, I think I am a good ESP teacher and feel efficacious. However, when the

students require other academic skills, I almost always evade their requests(participant 19).

Social Consequences

The second theme emerged in this study is categorized as the consequences of the tests on the society at large. The participants acknowledged that the entrance exams, particularly ESP sub-tests, certainly have great impacts on the society. The social impacts can be subcategorized as: deprivation from high education, injustice and unethical issues, and acceptance of unqualified candidates. Each of these subcategories is elaborated as follows .

- Deprivation from Education

The participants acknowledged that the results of selection tests- either norm-referenced or criterion-referenced- will certainly result in some candidates' deprivation from education. They believed that ESP tests, not unlike the other tests, do have such impacts. They believe the candidates whose English is good can achieve a good score on ESP tests, and are ranked higher than those with low English proficiency. Therefore, the ones with low English proficiency are deprived from studying for master and PhD degrees although their scores on the content sub-tests are not bad in comparison with the other candidates. One of the participants said:

I answered almost all items of the entrance exam except the ESP items. That is why, I did not pass the test but my friends who answered English items passed, and were accepted in very good and modern universities such as Tehran University. Two or three times I failed. Finally, I got tired, and did not attempt entrance examinations to post graduation schools. I was deprived. Such a failure is only due to my weakness in English (participant 20).

Another participant stated:

If I had been able to answer the items of the ESP sub-part of the PhD entrance examination, I would have passed the test,

and I could have become a highly educated person in my country. In fact, my only problem was my weakness in English. Damn English test turned out to be a disaster in my life. As a result, the society disadvantages the qualified people (participant 18).

- Injustice and Unethical Issues

The participants of the study also stated that the ESP tests administered at our universities lead to injustice and unethical issues. That is, almost all candidates acknowledged that Iranian people in different cities do not have access to the same educational facilities such as language and test preparation institutes to learn a foreign language, yet, they have to take the same norm-referenced tests. Therefore, any decision made based on the candidates' performance on these tests in which the candidates do not have the same privileges, is to a great extent unethical and unfair. The following examples are given to illustrate this theme:

In fact, I do not agree with the policy of test development and administration which is currently practiced in our country. Some of the test takers live in big cities with enough educational facilities. They can attend language classes. They can prepare instructional materials easily. They benefit from very experienced language teachers. I do not have the chance to make use of these necessary things. In Iran, the PhD candidates have to take the same test. Those who began learning English at the earlier age and attended different language classes can certainly answer all language test items. Sometimes, the candidates' scores on content courses are the same but their scores on English test are different. Therefore, I think it is not fair and ethical to make a decision about the candidate based on their differences in English scores (participant 12).

Another participant added:

It is really ridiculous. I know some of the master students whose scores on content sub-parts of the test such as applied

Chemistry, Physics chemistry were in fact below my scores. His score on ESP test was 90 but mine was 40. He was accepted but I was not. Do you think it is fair? I am sure that neither can he write a paper nor understand a lecture in English. So why should he pass but I fail? Really, it is unethical (participant 5).

The results of the study also indicate that although ESP tests play important roles in acceptance or non acceptance of the candidates into post graduate schools; ,no one knows for sure that these tests measure the learners' true ESP knowledge. Therefore, lack of correspondence between ESP test contents and target language use tasks will cause a kind of unethical and unfair issue. A participant mentioned:

How does one know that ESP tests are authentic, reliable, and valid? Certainly those with good scores are accepted. Even one point is important. But are those who scored high on ESP tests able to use language in target language use situations? ---- I really doubt. Is it fair to accept candidates based on the results of such unimportant tests? --- These are not fair and ethical (participant 19).

- Acceptance of Unqualified Candidates

Another emerged theme is the acceptance of unqualified candidates. The participants of the study believed that entrance examinations to tertiary and post graduate schools are all norm-referenced. In norm-referenced tests, even one decimal is determining. Naturally, the average score and percentile rank of the test takers are the criteria for acceptance or non- acceptance. Therefore, the ESP part of the entrance examinations is of much significance. Assuming that the candidates are all equal in their scores on content tests but different from each other in ESP tests, those with a better score on the ESP test are accepted. There are occasions when the more qualified candidates are rejected, and less proficient ones are accepted. These so- called PhD candidates will be managers, professors, etc. in future. They may be less qualified than the ones not accepted. The following examples are given to

illustrate this theme.

I think the candidates should be accepted or rejected just based on their scores on technical tests. How proficient one is in English is not important. Those who are good at technical subjects are in fact more qualified to enter post graduate studies than those who are proficient in English but weak at technical subjects (participant 16).

Another participant stated:

If the candidates are accepted based on their mean scores on technical and ESP sub-tests of the selection test, it is more likely that a good score on a separate sub-test influences the mean score, and increases the probability of the acceptance. The ESP sub-test can also have the same role. Therefore, those candidates whose mean score is positively influenced by the ESP score may be less qualified than those with proper scores on technical sub-tests but bad score on the ESP test (participant 11).

- Financial Consequences

The results of the study indicated that language tests particularly EAP entrance examination tests have great financial consequence for both individuals and parents. The students who are serious about entering a highly ranked university spend evenings, weekends, and even vacations preparing for the test at various exam preparation schools which provide a variety of coaching services. The participants acknowledged that supplemental education of this kind costs a good deal of money, and that the students and their families are willing to make such sacrifices. The financial consequences include textbooks, language institution, private tutors, and test preparation classroom expenses. The following examples illustrate this theme.

I am a student and I do not have too much money. I spent quite a lot of money on buying textbooks and sample EAP tests. I also spent some money for test preparation classes. Even a penny was important to me but I had to spend it to buy the needed books (participant 6).

Another participant stated:

My English was not good. I decided to go to some language teachers to teach me. The private tuition was somehow high. Although paying that amount of money was really difficult, I had to pay it (participant 14).

Entrance examinations, particularly language tests, sometimes have indirect financial consequences. The participants argued that instead of wasting time to learn English, they can work somewhere and earn a great deal of money. One participant argued:

I have MSc in Chemistry. I can have a good job with a great salary but I determined to get prepared for PhD entrance examination. I studied for about 12 months. If I had worked 12 months, I would have earned about 12000 dollars. I just studied hard but I failed the test only because of my bad performance in the English tests. My score on the other subjects was not bad. I am sure if I had got a better score on English subpart of the test, I could have passed the test successfully (participant 9).

- Family Consequences

The EAP/ESP tests have also some family consequences. The participants acknowledged that the results of these tests influence the family members of the test takers to a great extent directly or indirectly. They are emotionally, financially, and psychologically influenced. The following examples are given to illustrate this theme.

As I had to study hard for the entrance examinations, I had no more time to spend with my family (my wife and children). Whenever they asked me to take them out, I did not agree, because I just wanted to study. They got worried and they always complained. Two or three times my wife decided to divorce (participant 3).

Another participant stated:

I am married and I have to spend a part of my time with my family. But because of the importance of examination, I just

studied. When I was studying I could not earn enough money. Therefore, I could not meet my family's financial needs. It is their right to have everything they like. They sometimes did not understand me, and in fact, felt depressed (participant 13).

The test takers' psychological and social problems, caused by tests, indirectly influence their family members. One of the family members stated:

Iranian families are very emotionally closed. They cannot be indifferent to each other. Whenever a family member feels worried, depressed, or disappointed, the other members have the same feelings. When I was preparing myself for the test, I had to go to bed late and my family stayed awake (participant 20).

Discussion

The results obtained from a test can have serious consequences for individuals as well the programs, because many important decisions are made on the basis of tests results (Herman & Golan, 1993). Language learners and other participants may be influenced by official information about a test prior to its administration including advertising materials from the test publishers or by folk-knowledge such as reports from the students who have taken the tests earlier. They may also be affected by several sources of feedback following test administration. These would include the actual test scores provided by the exam scoring service, feedback from the test takers such as what was easy or difficult, what seemed fair or unfair, expected, or unexpected, feedback from the proctors, and feedback from the teachers in reaction to the students' scores (Baily, 1999). Taylor (2005) also believes that language tests can have consequences beyond just the classroom. Tests and test results have a significant impact on the career or life chances of individual test takers.

The impacts of different types of tests in different areas of the world have been studied empirically. The types of tests include national school examinations in Sri Lanka (Wall and Alderson,

1993; Wall, 1997,2000) , Israel Shohamy et al., 1996), and Hong Kong (Cheng, 1997, 1998); , university entrance examinations in Japan (Watanabe, 1997& 2004); , and international proficiency tests (Alderson and Hamp-Lyons, 1996). Much of what was revealed by these studies had to do with what Hughes (1989,1988) would call the ‘processes’ of teaching: the selection of content (skills, teaching materials, exam preparation materials), the methodology teachers used, and the ways in which they assessed their own students. The findings relating to participants often had to do with the stress and anxiety felt by teachers and learners.

In line with findings of the above mentioned impact studies, the results of the present study indicated that master and doctoral ESP tests administered as national entrance examination to Iranian state universities influence learners , ESP teachers, and society in different ways. The first theme emerged from the content analysis of the interviews was described as the psychological consequences. The psychological impacts were subcategorized into stress and anxiety, depression, disappointment, false self-confidence, and teaching efficacy. The findings of the impact studies carried out in different parts of the world only confirmed learners’ and teachers’ stress and anxiety before and after test administration whereas the results of the present study indicate that in addition to stress and anxiety, the learners become depressed and disappointed. Moreover, their self-confidence is negatively influenced by such tests.

Pearson (1988) says it is accepted that public examinations influence the attitudes, behaviors, and motivations of the learners, parents, and teachers. This influence is often seen as negative. The review of literature also indicated that examinations distort curriculum. The findings of the present study also indicated that a main negative consequence of ESP tests on Iranian ESP teachers is distortion of curriculum. That is, they ignore language skills which do not contribute directly to passing the exams. Rather, they just lay emphasis on teaching technical vocabulary and reading through limited teaching strategies such as translation to students’ native language. As such a kind of teaching method does not need proficiency in the

other language skills, the teacher thinks they are very efficacious. That is why, they think they have a high teaching efficacy.

The participants of the study also reported that ESP/EAP tests have social consequences. Not unlike the findings of washback studies, the results of the study indicated that social consequences of ESP tests including deprivation from high education, unfairness, and acceptance of unqualified candidates are all negative. Therefore, in line with Davies (1997), Messick (1989,1994, 1996), Hamp-Lyons (1997a, 1997b, 1989, 2000, 1999), and McNamara(1999), who assert arguing for a professional morality among language testers to protect the profession's members and individuals from the misuse and abuse of the test, it could be argued that ESP tests in Iran are unfair, and violate ethics assumptions. Viewing the use of ESP tests as instruments of social policy and control, and considering their gate-keeping function (Spolsky, 1997, 1981, 1994), it could be argued that ESP tests in Iran will lead to acceptance and rejection of some candidates whose true ability we are not sure of . Consequently, the society does not benefit from the qualified candidates since some qualified candidates are deprived from education while it is their own civil and social right to be accepted in universities. ESP tests practiced at Iranian universities, to put in words of Shohamy (1997, 1993,1998, 2000), contain contents or employ methods which are not fair to all test-takers. As the results, uses of such tests which exercise control and manipulate stakeholders rather than providing information on the proficiency levels seem to be against ethics of language testing.

The financial impact of ESP tests on test takers was the third theme emerged from the content analysis of the present study data. Due to the impacts of the entrance examinations on the career or life chances of individual test takers, they have to spend a great deal of money for preparation courses, sample tests, and even private tuition. As the majority of the test takers are students or unemployed, it is somehow difficult for them to earn money. Therefore, their life is greatly influenced by the results of

the tests. Moreover, the test takers who are bread winners of the family cannot satisfy their families financially as they do not have time to work somewhere to earn money.

Furthermore, the results indicated that the test takers' family members such as parents, children, husbands, and wives, are all directly or indirectly influenced by the tests consequences. The findings of the other studies indicated test consequences on the learners' parents; whereas, this study showed that, in addition to learners' parents, their children, husbands, and wives were all influenced by psychological, social, and financial consequences of the tests. Such clear difference between the findings of this study and the other studies is deeply rooted in the differences between the cultural and social values of the learners. In Iran, family members are closely related, and feel sympathy with each other.

Almost all the consequences reported by the participants were negative. That is, neither do they result in great innovations in learning and teaching ESP nor do they have positive and beneficial consequences on the stake holders. The great negative consequences of ESP tests, according to Kiani, Akbari, and Alibakhshi (forthcoming) are due to their lack of directness and authenticity. They believe that authentic tests will certainly lead to great positive consequences. Another justification for such negative consequences is the purpose of ESP tests and the decisions which are made on the basis of the tests. Naturally the negative consequences of norm-referenced tests are more serious than criterion-referenced ones.

Conclusion

This study was an attempt to explore the consequences of ESP tests on test takers and teachers' life and society. The assumption is that valid tests have positive consequences on the stake holders, society, and educational systems, and invalid tests have negative consequences on all stake holders, and distort curriculum. The results of the present study showed that the consequences of the ESP tests were all negative. Therefore, it could be concluded that these tests lack consequential validity

which is a main component of construct validity (Messick,). Viewing such a fatal shortcoming in ESP tests, we suggest that great changes in the contents, purpose and decisions made on the basis of these tests results are needed so that we can make a change in teaching ESP at our local universities. That is, it is concluded that the invalid tests should not be used as the instruments which filter the entry and nonentry of some candidates to state universities. Moreover, any decision made on the basis of invalid tests is against the critical issues of fairness and ethics. Therefore, it is essential that the test developers try to develop more authentic and direct tests, because authentic and direct tests have good washback validity.

The Authors

Gholam Reza Kiani is associate professor of applied linguistics in Tarbiat Modares University.

Goudarz Alibakhshi is an assistant professor of applied linguistics in Yasouj University.

Ramin Akbari is an assistant professor of applied linguistics in Tarbiat Modares University

References

- Alderson, J. C. (2004). 'Foreword' in L. Cheng, Y. Watanabe, and A. Curtis (eds.). *Washback in language testing: Research contexts and methods* (pp ?). London: Lawrence Erlbaum.
- Alderson, J.C., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14 (2), 115–129
- Alderson, J.C., & Hamp-Lyons, L.(1996). TOFEL preparation courses: a study of washback. *Language Testing*, 13(3), 280-97.
- Andrews, S. (1994a). 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. (1994). Assessing spoken English in public examination-why and how? In J. Boyle and P. Falvey (Eds.), *English Language Testing in Hong Kong* (pp. 57–86). Hong Kong: Chinese University.
- Bachman, L. (1990). *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.
- Bachman, L, & and Palmer, A. (1996). *Language testing in practice*. Oxford: Oxford University Press
- Bailey, K. M. (1999). *Washback in language testing*. Educational testing service. Peicton: Newjersey
- 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. (ed.) (1997). Special issue: ethics in language testing. *Language Testing*,14.
- Foucault, M. (1979). *Discipline and punish: The birth of the prison*. [Tr. A. Sheridan.] New York: Vintage Books.
- Hamp-Lyons, L. (1989). Language testing and ethics. *Prospect*, 5, 7-15.
- Hamp-Lyons, L. (1997a). Ethics in language testing. In: Clapham, C. (Ed.), *The Encyclopedia 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.
- Hamp-Lyons, L. (1999). Ethics, fairness(es), and developments in language testing. In: Elder, C., Brown, A., Grove, L., Hill, K., Lumley, T., McNamara, T. (Eds.), *Experimenting with*

- Uncertainty: Essays in Honor of Alan Davies; Studies in Language Testing 11 (pp. 212-217). UCLES/Cambridge University Press, Cambridge.
- Hamp-Lyons, L. (2000). Social, professional and individual responsibility in language testing. *System*, 28, 579-591.
- Herman, J.L., & Golan, S. (1993). The effects of standardized testing on teaching and schools. *Educational Measurement: Issues and Practices*, 12 (4), 20-25.
- Hughes, A. (1988). Introducing a needs-based test of English language proficiency into an English-medium university in Turkey. In Hughes, A.(Ed.), *Testing English for university study. ELT Documents 127 (pp.134-46)*. Oxford: Modern English Publications.
- Hughes, A. (1989). *Testing for language teachers*. Cambridge: Cambridge University Press.
- Kiani, R.G., Akbari, R., & Alibakhshi, G. (In press). On the authenticity of ESAP test. *Asian Journal of ESP*.
- Luxia, Q. (2005). Stakeholders' conflicting aims undermine the washback function of a high-stakes test. *Language Testing*, , 22, 142-173
- Madaus, G. (1990). Testing as social technology. Paper presented at the inaugural annual Boisi Lecture in Education and Public Policy. Boston College.
- Marvdast, A. B. (2004). *Qualitative research in sociology*. London: SAGE publications.
- McNamara, T. (1999). Validity in Language Testing: The Challenge of Sam Messick's Legacy. Messick Memorial Lecture, Language Testing Research Colloquium,
- Messick, S. (1989). Validity. In: Linn, R.L. (Ed.), *Educational Measurement (3rd Edition)(pp.13-103)*. American Council on Education/Macmillan, New York.
- Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, 23, 13-23.
- 1996. Validity and washback in language testing. *Language Testing*, 13, 241-56.

- 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). EL T Documents 128. London: Modern English Publications in association with the British Council.
- Shohamy, E. (1993). The power of tests: the impact of language tests on teaching and learning. NFLC Occasional Papers.
- Shohamy, E. (1997). Critical Language Testing and Beyond. Plenary address at the American Association of Applied Linguistics Conference.
- Shohamy, E. (1998). Critical language testing and beyond. *Studies in Educational Evaluation*, 24 (4), 331-345.
- Shohamy, E. (2001). *The Power of Tests*. London: Longman/Pearson.
- 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 Berry, R.(Ed.), *Bringing about change in language education: proceedings of the International Language in Education Conference* (pp. 55–66). Hong Kong: Department of Curriculum Studies.
- Spolsky, B. (1997). The ethics of gate keeping tests: what have we learned in a hundred years? *Language Testing*, 14 (3), 242-7.
- Spolsky, B. (1981). Some ethical questions about language testing. In: Klein-Braley, C.,Stevenson, D. (Eds.), *Practice and problems in language testing* (pp.5-30). Verlag Peter D. Lang: Frankfurt.
- Strauss, A., & Corbin, J. (1990). *Basics of qualitative research: grounded theory procedures and techniques*. Newbury Park, CA: Sage.
- Taylor, L. (2005). Washback and impact. *ELT Journal*, 59(2).
- Wall, D., (1997). Impact and washback in language testing. In: Clapham, C., Corson, D. (Eds.), *Encyclopedia of Language and Education*, Volume 7: *Language Testing and Assessment* (pp. 291–302). Kluwer, Dordrecht.

- 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, 41–69.
- Watanabe, Y. (1997). Constructing a classroom observation scheme for the test impact research. *Sophia Linguistica*, 41, 297-313.
- Watanabe, Y. (2004). Methodology in washback studies. In L. Cheng, Y. Watanabe & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods*, (pp.19-36). Mahwah, N.J.: Lawrence Erlbaum