

## **Examining Washback of Language Assessment within an Educational Context: Voices from Teachers and Examiners**

Zia Tajeddin

*Associate Professor, Allameh Tabataba'i University*

Masoomeh Estaji

*PhD, Allameh Tabataba'i University*

### **Abstract**

Teaching enriches students' learning contents, while assessment evaluates students' learning results. In recent decades, assessment has gained increased attention in EFL education. One of the important issues in the practice of assessment is the washback effect of testing on teaching and learning (Alderson and Wall, 1993; Buck, 1988; Hughes, 2003). Washback effect on teaching and learning can be positive (beneficial) or negative (harmful). Much research has been focused on the washback effect of assessment on students' learning or large-scale, standardized tests (Watanabe, 1996). Assessment has powerful influence on teaching, too. Wall (1998) claimed that high-stakes tests might induce the impact on teaching methodology and content.

This study explored the washback effect (Alderson and Wall, 1993) of a high-stakes exam, TOEFL iBT. It focused on the teachers' perceptions of the test and its washback effect on teaching. The purpose of this qualitative study was to describe the washback behaviors of teachers and, to a lesser extent, students in the high-stakes testing environment of TOEFL iBT. The study followed a group of five teachers, teaching TOEFL iBT preparation courses. To explore the washback phenomenon, this study employed various methodological techniques, including questionnaire surveys, classroom observations, and in-depth interviews.

## 2 Examining Washback of Language Assessment...

The results of this study have shown a multi-layered account of the washback phenomenon. Teachers' different perceived levels of awareness of high-stakes exams and perceived students' learning attitudes have a crucial influence on teachers' perceptions of the impact of high-stakes exams on their curricular planning and instruction. However, several discrepant findings from this study further support that washback is quite context-oriented and complex. The results imply that simply examining one factor without a covariance analysis or examining the phenomenon in one context is not capable of explaining critical washback issues, such as how and why washback phenomenon influences some teachers but not others. It is recommended that longitudinal studies, such as long-term classroom observations, should be conducted in order to explain to what extent washback actually occurs to influence classroom teaching.

**Key words:** Washback, Impact, Test Consequences, Consequential Validity, Test Usefulness, Teacher Washback, Assessment Usefulness

**1. Introduction**

Testing tends to induce consequences for its stakeholders. It is well known in the field of education that there is a set of relationships, intended and unintended, positive and negative, between testing, teaching, and learning. Alderson and Wall (1993) suggested that the term "washback" provides a useful metaphor to help us explore the role of language tests in teaching and learning, i.e. in relation to factors such as the individual learner, the teacher's attitudes and behavior, the classroom environment, and the choice and use of teaching/learning materials. "Washback" allows for the possibility of effects of tests on teaching to be viewed on a continuum – stretching from negative (harmful) at one end, through neutral, and into positive (beneficial) at the other end. Negative washback is said to occur when test content or format is based on a narrow definition of language ability and so constrains the teaching/learning context. Positive washback is said to result when a testing procedure

encourages “good” teaching practice and positive learning outcomes.

However, the functions of tests in reality are far more beyond the intrinsic role as evaluation instrument, or the practical part as information resources for pedagogical refinement. As Davies notes, a test is “so potent in influence, so salient a presence, deserves much closer attention and study than it typically receives” (1990, p. 1). The power of examination over what takes place in the classroom has been widely reported by many educationalists. Tests may, in essence, change people’s perspectives on what is occurring in the social context as well as in teaching, or influence their behaviors. The power of tests is seen primarily in their ability to shape behavior. A test may even make one’s life different in some social context. Spolsky holds that “a test can help some people but cause harm when its results are misapplied or its aims distorted” (1999, p. 6). The belief that assessment can leverage educational change has often led to top-down educational reform strategies in order to bring about changes in teaching and learning by bringing about changes in testing.

However, such a reform strategy was counter-argued by Andrews (1994) as a “blunt instrument” for bringing about changes in teaching and learning. The actual teaching and learning situation is clearly far more complex. Each different educational context (school environment, messages from administration, expectations of other teachers, and students) plays a key role in facilitating or detracting from the possibility of change.

The focus of the study is on the perceptions of teachers involved in the TOEFL iBT courses. One reason for this is that studies indicate that the attitudes and perceptions of the candidates are highly influenced by the teacher (Spratt, 2005). It is natural, and in many cases appropriate, that students are somewhat apprehensive of a benchmark exam- but it is the teachers’ attitudes to a particular exam, and the range of activities that they use to prepare for it, that will be a major determiner of the balance of positive and negative washback in the classroom.

### 1.1 The Influence of Tests on Instruction: Washback

Deliberate actions focusing on the communicative abilities that students are expected to develop, and that teachers are expected to facilitate in and out of the classroom, have generated the belief that the new means used to assess skills (authentic or performance based tests) can offer teachers more accurate opportunities to measure student abilities. These new means of testing and assessment can include, but are not limited to, oral exams, letter writing, descriptions of pictures or story telling, debates, group projects such as newscasts, publishing a newspaper, or skits depicting different situations such as visits to the doctor's office, the grocery store, the airport, etc. Typically these activities also reflect the national *Standards for Foreign Language Learning*. Because these assessments are more and more commonly built into the foreign language curriculum, they become part of daily class activities in an attempt to prepare students for the different types of tasks and activities that they will be asked to complete. As a result, tests and assessments of this nature are generally seen to render a positive influence on classroom instruction and practice (Valette, 1994; Wiggins, 1994. The connection between tests and the methods and instructional techniques employed in the classroom is referred to as washback. Washback is "...the influence of language testing on teaching and learning" (Cheng, Watanabe, and Curtis, 2004, p. 13). Messick elaborates:

... in the case of language testing, the assessment should include authentic and direct samples of the communicative behaviors of listening, speaking, reading, and writing of the language being learnt. Ideally, the move from learning exercises to test exercises should be seamless. As a result, for optimal positive washback there should be little, if any, difference between activities involved in learning the language and activities involved in preparing for the test. (Messick, 1996, p. 241-2)

Messick refers to "optimal positive washback," which given the link between the activities that students perform in the classroom, and the task and performances that they are asked to

perform for assessments, instruction and testing or assessment, would appear to operate in complete synchronization. However, this is not always the case due to a variety of factors, the most notable in present educational settings being the importance given to national and state standardized exams, or district or school wide exams required to meet graduation requirements. The importance given to these exams can often influence teaching and instruction in negative ways. In other words, not all washback is positive. In particular, many language teaching specialists worry about the practice of “teaching to the test,” in which case sacrifices are made in what is taught or how it is taught in order to ensure that students can produce the type of language required on a test. This is referred to as negative washback. Negative washback has been observed and documented through such activities as teachers ceasing to use certain materials and activities that are not included on mandated tests, teachers including specific activities and or materials that are known to be on the tests, and giving mock tests to students using classroom time to administer these tests and discuss answers (Alderson and Hamp-Lyons, 1996; Shohamy, Donitsa-Schmidt, and Ferman, 1996).

It is undeniable to anyone who has ever experienced a test that instruction is usually influenced to some degree by a test. Questions such as “Do we have to know this for the test?”, “Is this going to be on the test?”, “How should I study for the test?”, “How will I know which answer you will want on the test?”, permeate all classrooms in every subject area. The age of the student does not matter; the bottom line is that everyone wants to be able to prepare for a test, and that means that it is necessary to know what is on a test. Because teachers have taken and given many tests, they are in a position to adapt their behavior and their teaching techniques to provide instruction and materials that will help to prepare students for the demands of the test. This is washback in its most basic form, often unconsciously appearing in daily class sessions.

### 1.2 Teachers and Washback

Besides learners, teachers are the second group that are directly affected by high-stakes tests because tests will influence teaching, will influence what and how teachers teach, the rate and success of teaching, degree and depth of teaching and will influence attitudes to content and method of teaching and learning (Alderson and Wall, 1993). The importance of teachers in washback processes is emphasized by Alderson and Wall (1993, pp. 120-121) in several of their restatements of the washback hypothesis:

1. A test will influence teaching.
2. A test will influence what teachers teach.
3. A test will influence how teachers teach.
4. A test will influence the rate and sequence of teaching.
5. A test will influence the degree and depth of teaching.
6. A test will influence attitudes to the content, method, etc. of teaching and learning.

They also pose two contradictory hypotheses which demand investigation: (1) “tests will have washback on all learners and teachers,” and (2) “tests will have washback effects for some learners and some teachers, but not for others.”

The vast majority of the available empirical research on washback includes at least some focus on teachers. In fact, it is safe to say that teachers are the most frequently studied of all the participants in the washback process.

Bachman and Palmer (1996) claim that teachers may find teaching to the test almost unavoidable although they may personally prefer to teach certain material in a specific way. They add that the term “teaching to test” implies doing something in teaching which may not be compatible with teachers’ own beliefs, values and goals.

Several studies are done on the impact of tests on educational systems and language teaching methods (Cheng, 1997; Hughes 1988; Lam, 1994; Qi, 2005). It is a generally accepted fact that teachers do accommodate their teaching based on the exam that their students are going to take. Cheng gives an example of a

change in Hong Kong Certificate of Education examination and reports that teachers gave up practicing reading aloud type of activities after the test excluded this part and instead they started role-play tasks and group discussions which took place of reading aloud. However, this is not the case all the time. A change in the test does not always guarantee a beneficial change in the language curriculum. Qi reports about a change in National Matriculation English Test in China which had two major purposes; selecting students and affecting teaching and learning favorably. It has been found out that these two purposes were conflicting with each other so rather than changing the curriculum toward a communicative one, they ended up with the same kind of memorization of grammar rules and vocabulary.

In most of the studies, it is seen that the trials of the policy makers in changing a test so that it will end up changing in actual teaching practice were not successful. So, instead of behaving tests as servants to teaching, they should be seen as co-workers. Test designers, teachers and even test takers should be involved in the process of test preparation.

### **1.3 Washback Research**

The most commonly held view of washback is perhaps best expressed by Alderson in his comments on his own research in the area of language washback as it began a decade ago:

I believe there is no longer any doubt that washback does indeed exist. ...The question today is not "does washback exist?" but much rather what does washback look like? What brings washback about? Why does washback exist? (2004, p. 9)

Investigating early washback, Wall and Alderson conducted a landmark study of washback in language testing in Sri Lanka from the middle of 1988 to the end of 1991. The purpose was to investigate the impact of a new "O-level" English examination introduced in the Sri Lankan educational system in 1988. The importance of English in Sri Lanka is great, as many depend upon it to conduct international business, national business, and for social

purposes. The “O-level” examination, administered during what corresponds to North American 11th grade, is an exam that students must take upon culmination of their 11th year of study. A student’s grade on this examination determines whether or not s/he will be allowed to continue with further study; or, if s/he enters into the workforce, if s/he will be suitable for sought-after jobs. In an effort to have students develop more practical English skills, the Ministry of Education developed new textbooks that placed a greater emphasis on reading and writing with a purpose, and on oral skills. In order to have these new innovations “taken seriously,” it was determined that a new “O-level” test would accompany the textbook series to force the teachers and the students to meet the demands of the test, and consequently improve instruction and achievement (1996, p. 197).

The study included an evaluation of the O-level exam, focusing on validity and reliability of the test, and the impact that the test had on the classroom. The research was conducted in conjunction with Lancaster University (Wall and Alderson, 1996). In this study the researchers developed the basic framework for the *Washback Hypotheses*, which simply stated, posited that tests influence teaching. The specific hypotheses as developed by Alderson and Wall (1993, p. 120-121) are as follows:

1. A test will influence teaching.
2. A test will influence learning.
3. A test will influence what teachers teach.
4. A test will influence how teachers teach.
5. A test will influence what learners learn.
6. A test will influence how learners learn.
7. A test will influence the rate and sequence of teaching.
8. A test will influence the rate and sequence of learning.
9. A test will influence the degree and depth of teaching.
10. A test will influence the degree and depth of learning.
11. A test will influence the attitudes to content, method, etc. of teaching/learning.
12. Tests that have important consequences will have washback.



13. Tests that do not have important consequences will have no washback.
14. Test will have washback on all learners and teachers.
15. Tests will have washback effects for some teachers and some learners, but not for others.

Attempts to document washback with empirical evidence have resulted in existing washback research that has been qualitative in nature. Qualitative research differs from experimental research or survey research in that the researcher is interested in “describing in detail all of what goes on in a particular activity or situation rather than on comparing the effects of a particular treatment..., or describing the attitudes or behaviors of people...” (Fraenkel and Wallen, 2003, p. 430).

Table 1 summarises the methods and findings of recent case study investigations of washback in language education that have included an observational element. These studies covered a wide range of educational contexts, with observation either focusing on a small number of participants observed intensively over a sustained period (Alderson and Hamp-Lyons, 1996; Read and Hayes, 2003) or on a lighter sampling of classes to allow for observation of larger numbers of teachers and a broader perspective (Hawkey, 2006; Wall, 2005).

Testing innovations are often designed to exploit the supposed potential of tests to encourage learning of targeted skills. Shohamy, Donitsa-Schmidt, and Ferman (1996) document how a new test of oral communication was introduced to encourage the teaching of speaking skills.

However, research into washback has consistently shown that tests are not, of themselves, necessarily effective as “levers for change” (Pearson, 1988); successful educational innovations require both concerted system-wide reform and extensive support from those affected (Wall, 2000). It is now clear that washback involves complex interactions between tests, teachers and learners, which determine whether individuals will embrace or reject intended change.

**Table 1: Studies of Washback in Language Education**

Study	Exam Studied	Teaching/ Learning Context	Main Issues Addressed
Alderson and Hamp-Lyons (1996)	TOEFL	Specialized language institute in US	The influence of TOEFL on classroom teaching
Brown (1998)	IELTS	University-based language institute	The influence of IELTS preparation in improving IELTS scores from entry to exit
Burrows (1998, 2004)	Certificates Spoken and Written English (CSWE)	Adult English Migrant English Program in Australia	Links washback to theories of change
Cheng (2005)	Hong Kong Certificate Examinations in English (HKCEE)	3 HK secondary schools (main study)	Differences in how teachers cope with change Extensive support required in implementation of innovation
Cheng (1997, revised 1998)	(HKCEE) in 1994	Hong Kong secondary schools	The possible washback effect from the exam on the teaching of English in Hong Kong secondary schools
Hamp Lyons (1998)	TOEFL	TOEFL studies	The role of textbooks in test washback
Hawkey (2006)	IELTS	10 language schools in UK, Japan, Cambodia	Teacher preference for task-based activities, involving micro skills relevant to IELTS
Read and Hayes (2003); Hayes and Read (2004)	IELTS	2 language schools in New Zealand	Great pressure to "teach to the test" private language schools
Shohamy et al. (1996)	English Foreign Language Test (Arabic Second Language Test)	Secondary schools in Israel	Impact of tests on classroom activities, time allotment, teaching materials, prestige of subject tested, promoting learning Views of different stakeholders: language inspectors, teachers and students
Turner (2001)	Exams in English As a Second Language	Canadian French-speaking primary and Secondary schools	Development of rating scales and its consequential effects on the teachers involved in the development
Wall (2005)	Sri Lankan 'O' level	50 secondary schools in 11 areas of Sri Lanka	Neglect of speaking skills traceable to exam content
Watanabe (1996, 2004)	Various Japanese university entrance examinations	3 High schools in Japan	Teachers vary in their approaches to exam preparation _ informed by attitudes toward examinations

#### **1.4 Washback as a Criterion for Developing and Evaluating Language Tests**

Positive washback, by whatever name, has recently been recognized as one of the main criteria for evaluating language tests. In his 1979 book, *Language Tests at School*, Oller identifies the key characteristics of a good test as being reliability, validity, practicality (also called “feasibility”), and instructional value- the last being most closely related to current conceptions of washback.

Washback has received even more attention as an evaluative criterion recently, with the advent of communicative language testing. For instance, one of Morrow's (1991) five criteria for evaluating communicative language tests is the idea that such tests should “reflect and encourage good classroom practice” (p. 111). In describing a test development project called the Communicative Use of English as a Foreign Language, Morrow states: “This (i.e., washback) is a major concern underlying the design of tests; indeed in many ways the tests themselves have drawn on ‘good’ classroom practice in an attempt to disseminate this to other classrooms” . Morrow says that a “conscious feedback loop between teaching and testing, in terms of not only content but of approach, is a vital mechanism for educational development”(see also Shohamy, 1993).

Boyle and Falvey (1994) observe that “there has been a recent renewal of interest in the link between good teaching and good testing” (p. 11). They also note that washback, along with validity, reliability and practicality is now “one of the Big Four considerations in evaluating the worth of a test”.

#### **1.5 Purpose of the Study**

Research in the field of language testing indicates that washback is a highly complex phenomenon rather than a unitary notion. Amongst various findings, relatively well established is that washback is a function of the test and other factors. However, This study intends to explore the nature and scope of the washback effect on aspects of teachers’ perceptions and teachers’ behaviors within the context of TOEFL iBT and non-TOEFL courses.

Questions addressed in this study include:

1. Does the nature and scope of washback have any effect on teachers' perceptions of aspects of teaching toward TOEFL iBT?
2. Do the teachers working on TOEFL iBT courses have a different perception from those working on non-TOEFL course?

## **2. Method**

A combined research framework, using multiple approaches, was employed in this study in order to explore a clear and convincing picture of the washback phenomenon.

### **2.1 Participants**

The sample of the study included teachers at MFT (Mojtame Fanni Tehran). Based on the nature and the methodological considerations of this study, the teachers were selected using purposeful sampling (Patton, 1987), certain criteria, and the main purpose was to select teachers based on whether they could provide a rich variety of information about classroom teaching and learning activities in relation to TOEFL iBT.

The teacher participants included 5 individuals working on non-TOEFL English courses in advanced levels and 5 on TOEFL iBT preparation courses. Although teacher participants were drawn from the same institution, the teachers were not necessarily from the same classes. However, teachers were working with the same course outlines and questionnaire responses from those on courses of the same type indicated that their practices were consistent. This consistency between teachers was felt to justify the comparison of teacher responses that follows.

### **2.2 Instruments**

Two types of instruments were used:

- 1. Teacher Questionnaires:** This paper takes as its focus a section of teacher questionnaires that teachers completed. The teacher questionnaire was designed to obtain preliminary data on teachers' reactions toward TOEFL iBT, their perceptions and understanding of the examination, and what they would like to do to prepare their students for the examination. The questionnaire was designed on a multiple-choice format 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 classes.

2. **Teacher Interviews:** Teacher interviews were conducted and these were followed up with individual interviews. As many as 8 teachers were selected. Interviews were conducted with the teachers to gather information regarding their attitudes toward the TOEFL iBT and changes in their strategies before and during test preparation. In other words, the interview questions were designed to explore teacher beliefs: the interviews examined whether teachers believed that their teaching had been influenced by the introduction of TOEFL iBT; and whether they reported changes to their teaching practices. These questions were based on three washback hypotheses taken from Alderson and Wall (1993): "A test will influence what teachers teach"; "A test will influence how teachers teach;" and "Tests will have washback effects for some teachers, but not for others" (pp. 120-121). The interviews were then transcribed and analyzed.

#### **Data Collection and Analysis Procedures**

The teachers were asked to complete the questionnaire to collect information about their beliefs, attitudes, and their perceptions of TOEFL iBT, and to record any changes in their perceptions of the test. The data were then collected and analyzed. The data gathered and analyzed for this study that pertains directly to the teachers in the study is reported in a "multiple- or collective case study" format (Fraenkel and Wallen, 2003, p. 439). The result allowed the researcher to make comparisons between cases (the teachers of TOEFL iBT and non-TOEFL courses), even in instances where case comparisons appeared to result in contradictions.

The interviews with the teachers were held on a one-to-one basis. Each interview lasted for about one hour. All interviews were recorded and then transcribed. All interviews were recorded on tape, which were transcribed and reviewed later on the same day.

Following the data collection phase, responses were then summarized and compared with responses of other teachers.

### 3. Results and Discussion

In the following sections, the discussion will follow the 5 questions of the questionnaire and present the statistical results in a table. This is followed by a discussion of the findings.

#### 3.1 Teachers' Reactions to TOEFL iBT

When teachers were asked about their reactions to TOEFL iBT, there was a significant difference in their reactions. The results are presented in Table 2.

**Table 2:** Teachers' Reactions to TOEFL iBT

Teachers' reactions	Teachers working on TOEFL iBT courses N=5	Teachers working on non-TOEFL courses N=5
Skeptical about the change	0	3
Neutral about the change	0	1
Welcome the change	4	1
Enthusiastically endorse the change	1	0

The great majority of people teaching TOEFL iBT courses "welcome the change," whereas teachers of non-TOEFL courses were "skeptical about the change." TOEFL iBT teachers "enthusiastically endorse the change." This is a clear indication of teachers' positive attitudes toward TOEFL iBT. This positive attitudinal change in teachers toward TOEFL iBT is more observed among TOEFL iBT teachers than those who are involved with advanced-courses and presumably skeptical about the change. Teachers of exam preparation courses tend to have a positive and supportive attitude toward change in general. It is only when they come across problems and difficulties in actual teaching that they

confront the pressure of change. The reluctant or skeptical attitude toward change might be due to the practical aspects of teaching in the case of this examination change. Moreover, teachers seldom abandon what they have been doing and embrace completely some new philosophy, methodology, or new curriculum approaches, even if they do have a positive attitude toward the examination change. It is only to be expected that teachers would modify what they have been doing to prepare their students for TOEFL iBT exam as it is so important to both students and teachers alike. However, even if the survey showed a positive attitudinal change, this does not necessarily mean that the teachers are going to change their behavior.

### **3.2 Teaching Planning and Medium of Instruction**

There were two categories in this section. The results are shown in Table 3, and each category will be discussed in detail below. Among the two categories in this section, the first category, “how teachers arrange their teaching in classes” showed TOEFL iBT teachers arrange their teaching according to the integration of skills. In addition, they seemed to arrange their lessons much more according to “separate skills” and “content to be taught” than according to “textbooks” and “language activities”. Of the other two choices, “textbook” and “language tasks” remained relatively unchanged. In contrast, non-TOEFL teachers arranged their lessons much more according to “textbooks” and “separate skills.” Furthermore, they seemed to arrange their lessons much more according to “integration of skills” than according to “the content” and “language activities.”

**Table 3:** Teaching Planning and Medium of Instruction

<b>Items</b>	<b>Variables</b>	<b>Teachers working on TOEFL iBT courses N=5</b>	<b>Teachers working on non-TOEFL courses N=5</b>
How do teachers arrange their teaching in classes?	According to textbooks	0	2
	According to integration of skills	3	1
	According to separate skills	1	2
	According to the content	1	0
	According to language activities	0	0
What is the medium of instruction	English only	4	3
	English supplemented with occasional Persian explanation	1	2
	Half English and Persian	0	0
	Mainly Persian	0	0

However, the item regarding medium of instruction showed an interesting pattern. Teachers in TOEFL iBT classes used, to a large extent only English as their medium of instruction whereas teachers of non-TOEFL courses used English and English supplemented with occasional Persian explanation. This seems to indicate that TOEFL iBT teachers tended to use English only more often, with one possible reason for this being their concerns and worries over their current students' English levels. Teachers focused on the target



language so that students could pass the exam, which was frequently mentioned by many TOEFL iBT teachers during the interview. It might suggest that the washback effect of TOEFL iBT was also observed on general teaching planning and preparation.

**Textbook Arrangements Related to Teaching Materials**

There were two categories related to teaching materials in the context of TOEFL iBT. They were presented in a multiple-choice format as shown in Table 4.

From Table 4, it can be drawn that teachers of both TOEFL iBT and teachers of non-TOEFL courses have much say in the choice of textbooks. Together with the panel chairs, it is normally the teachers of English who decide which textbooks to use in their teaching. As one of the panel chairs explained, it was natural and essential to choose the textbook which teachers preferred; otherwise they would not enjoy teaching it.

**Table 4:** Textbook Arrangements

Items	Variables	Teachers working on TOEFL iBT courses N=5	Teachers working on non-TOEFL courses N=5
Who makes the major decision on the choice of textbooks?	1. Panel chair	2	2
	2. Panel of experts	1	1
	3. English teachers together	2	2
	4. Yourself	0	0
What are the primary functions of textbooks in teaching?	1. To provide practical activities	1	1
	2. To provide a structured language program to follow	2	2
	3. To provide language models	1	1
	4. To provide information about the language	1	1

Regarding the function of the textbook in teaching, it can be drawn that textbooks for both TOEFL iBT and non-TOEFL teachers, are “to provide a structured language program to follow.” This indicates that textbooks play an important role in English teaching.

The findings from the teacher survey illustrate further the complex nature of washback effects. In exploring teachers’ reactions to TOEFL iBT, TOEFL iBT teachers reacted positively to this examination change. However, the most important aspects that governed teachers’ teaching TOEFL iBT courses remained relatively unchanged. Teachers were examination oriented, and teaching was content-based, integrated, and highly controlled. It is also interesting to note that the studies that found evidence of washback on teaching also found large differences in the way teachers teach toward the same exam, with some adopting much more overt “teaching to the test”, while others follow more creative and independent approaches.

### **Discussion**

Major issues explored in this study were as follows:

First, the survey showed that teachers had a positive reaction toward TOEFL iBT. For example, fewer teachers tended to be skeptical about the change. The majority of teachers welcomed the change. Teachers seemed to have a positive attitude toward the change, but the results also suggest a reluctant attitude toward making the changes that they ought to carry out in their own teaching.

Second, in summarizing the results of how the teaching materials are decided upon for TOEFL iBT, it appears that teachers had a strong voice in the choice of textbooks. According to teachers, the major function of the textbook was to provide a structured language program to follow in their teaching. The results indicated the important function that textbooks played in the teaching of English in both TOEFL iBT and non-TOEFL classrooms.

Third, summarizing the findings of washback on classroom teaching behaviors, it can be seen that the medium of instruction remained unchanged despite the examination change. There was also a tendency for teachers to pay more attention to the content, skills to be taught, and practice tests to be given to students. This might indicate that teachers paid increased attention to the examination. Some teachers mentioned that they assigned more homework to prepare their students for TOEFL iBT. The results also showed that teachers paid more attention to teaching content than teaching methods and other teaching and learning factors in their lesson preparation.

Summarizing the above aspects of teaching in relation to TOEFL iBT, the changes produced by TOEFL iBT were seen to be superficial rather than substantial. It appears that changing the examination had likely changed the kind of exam practice, but not the fact of the examination practice. Changing an examination's format does not tend to change the degree of emphasis on the examination. Furthermore, from the findings, it can be seen that teachers were examination oriented, and their teaching was content-based and highly controlled by the teachers themselves. It is important to point out that the short period of research time might have restricted the scope of the research findings so that only the superficial changes were observed.

#### **4. Conclusions and Implications**

The most significant impact studies have been conducted in various countries in high- stakes testing environments, as these are places where one finds the most common use of large-scale language tests.

The principal focus of the previously cited studies, as well as this study, is washback in foreign language classes. However, in an era when high-stakes testing is gradually becoming the law of the land in the United States, as dictated by the federal No Child Left Behind legislation, the research reported here has important implications for a wide range of disciplines beyond foreign languages because of the prevalence of testing. As a result of recent and ongoing education reform in the United States, it appears that

policy makers will continue to view assessment, particularly in the form of high-stakes, standardized tests, as an appropriate method for spurring change in curricula and instruction. According to Shohamy, the appropriateness of this approach to school reform is debatable:

Tests...can open or close doors, provide or take away opportunities, and in general shape the lives of individuals in many different areas. ...tests are used as a method of imposing certain behaviors on those who are subject to them. Tests are capable of dictating to test takers what they need to know, and what they will learn and what they will be taught. The use of tests as disciplinary tools means that test takers are forced to change their behavior to suit the demands of the test. ...Using the tests as disciplinary tools is an extension of the manipulation of tests by those in authority- policy makers, principals, and teachers- into effective instruments for policy making. It is the realization that test takers will change their behavior in order to succeed on tests that leads those in authority to use tests to cause a change in behavior in accordance with certain priorities. (2001, pp. 16- 17)

Students in foreign language classes are becoming increasingly more diverse and are studying languages for a variety of reasons. The undergraduate, intermediate level classroom as the focus of this study provides a particular challenge for both students and educators, as this is the level where students are first expected to use the target language and refine their skills (Rava, 2000). The transition to this level can be problematic for both students and teachers, especially when expectations and educational objectives for each group are not always clearly articulated. As Rava (2000) explains, "faced with the challenge of striking a balance between institutional language requirements and preparation for more advanced language and literature studies, faculty and students alike ...often view the intermediate curriculum as the ugly stepsister" (p. 342). If foreign language education aims to meet the needs of students through language programs, and to prepare them for success in an increasingly global society, then educators must seek

to discover ways by which teaching and assessment can be linked as seamlessly as Messick (1996) describes. To do so, educators and policy makers must continue to increase the knowledge of washback in the classroom. Tests that are a regular part of the curriculum are considered low stakes tests, and virtually part of all classes in the majority of academic institutions in the United States. Consequently they provide an opportunity to study washback as it occurs in a variety of educational settings. As seen above, research to date has not been much successful in providing a clear and convincing picture of the extent to which and ways in which washback is manifested in foreign language instruction. Focusing on a high-stakes testing environment (a unified testing system in place) should help bridge this gap, especially since the tests and assessments used in the modern foreign language classroom focus more on the standards, and try to include a range of student skills and abilities. A focus on skills and abilities is also something that will be an important component of any national or standardized test that is developed and implemented for foreign languages. As a result, this paper can provide insight for other disciplines with respect to the influence that tests have on student behaviors and corresponding teachers' instructional behaviors and programs. Focusing on instruction can help to provide teachers with the critical perspective needed for improvement, as well as the impetus for change in instructional behaviors when needed. Without this type of instructional focus, teachers will continue to emphasize what they believe are the important aspects of language learning, whether or not they are based on skills, and/or included on the tests. Teacher beliefs influence teacher behaviors deeply no matter how the curriculum, course objectives, and the testing program work together. One recommendation for a future washback study would be to investigate the effect of washback on teachers of different teaching beliefs and different teaching styles using different methodologies.

In conclusion, Tests will never be eliminated from educational institutions; therefore, it is best to embrace them and their power. Tests can drive change, and if the intention is to make changes to

foreign language instruction, it must be done consciously. Efforts must be taken to help teachers encourage positive washback and reduce negative washback.

## References

- Alderson, J. C. (2004). Foreword. In L. Cheng, Y. Watanabe, and A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 9-12). Mahwah, NJ: Lawrence Erlbaum Associates.
- Alderson, J. C., and Hamp-Lyons, L. (1996). TOEFL preparation courses: A study of washback. *Language Testing*, 13(3), 280-297.
- Alderson, J. C., and Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14(2), 115-129.
- Andrews, S. (1994). The washback effect of examinations: Its impact upon curriculum innovation in English language teaching. *Curriculum Forum*, 1(4), 44-58.
- Bachman, L. F., and Palmer, A. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University Press.
- Boyle, J., and Falvey, F. (Eds.) (1994). *English language testing in Hong Kong*. Hong Kong: The Chinese University Press.
- Brown, J. D. H. (1998). An investigation into approaches to IELTS preparation, with particular focus on the Academic Writing component of the test. In S. Wood (Ed.), *IELTS Research Reports*, Vol.1, (pp. 20-37). Sydney: ELICOS/IELTS Australia.
- Buck, G. (1988). Testing listening comprehension in Japanese university entrance examinations. *JALT Journal*, 10(1), 15-42.
- Burrows, C. (1998). *Searching for washback: An investigation of the impact on teachers of the implementation into the adult migrant English program of the assessment of the certificates in spoken and written English*. Unpublished doctoral dissertation, Macquarie University- Sydney.

- Burrows, C. (2004). Washback in classroom-based assessment: A study of the washback effect in the Australian Adult Migrant English Program. In L. Cheng, Y. Watanabe, and A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 113-128). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Cheng, L. (1997). How does washback influence teaching? Implications for Hong Kong. *Language and Education*, 11(1), 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.
- Cheng, L. (2005). *Changing language teaching through language testing: A washback Study*. Cambridge, UK: Cambridge University Press.
- Cheng, L., Watanabe, Y., and Curtis, A. (Eds.) (2004). *Washback in language testing: Research contexts and methods*. London: Lawrence Erlbaum.
- Davies, A. (1990). *Principles of Language Testing*. Oxford: Blackwell.
- Fraenkel, J.R., and Wallen, N.E. (2003). *How to design and evaluate research in education*. New York: McGraw Hill Higher Education.
- Hamp-Lyons, L. (1998). Ethical test preparation practice: The case of the TOEFL. *TESOL Quarterly (Forum Commentary)*, 32(2), 329-337.
- Hawkey, R. (2006). *Studies in language testing 24: Impact theory and practice: Studies of the IELTS test and Progetto Lingue 2000*. Cambridge: Cambridge University Press and Cambridge ESOL.
- Hayes, B., & Read, J. (2004). IELTS test preparation in New Zealand: Preparing students for the IELTS Academic Module. In L. Cheng, Y. Watanabe & A. Curtis (Eds.), *Washback in language testing: Research context and methods* (pp. 97-112). Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.

- 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). London: Modern English Publications in association with the British Council.
- Hughes, A. (2003). *Testing for language teachers* (2nd Ed.). Cambridge: Cambridge University Press.
- Lam, H. P. (1994). Methodology washback \_an insider's view. In D. Nunan, R. Berry, And 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.
- Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13, 241-256.
- Morrow, K. (1991). Evaluating communicative tests. In S. Anivan, (Ed.), *Current Developments in Language Testing*, RELC Anthology Series, 23, RELC.
- Oller, J. W. (1979). *Language tests at school: A pragmatic approach*, London: Longman.
- Patton, M. Q. (1987). *How to use qualitative methods in evaluation*. Beverley Hills, CA: Sage Publications.
- Pearson, I. (1988). Tests as levers for change. In D. Chamberlain and R. J. Baumgardner (Eds.), *ESP in the classroom: Practice and evaluation. ELT Documents*, Volume 128 (pp. 98-107). London: Modern English Publications.
- Qi, L. (2005). Stakeholders' conflicting aims undermine the washback function of a high- stakes test. *Language Testing*, 22(2), 142-173.
- Rava, S. (2000). The changing face of the intermediate language curriculum. *Foreign Language Annals*, 33(3), 342-348.
- Read, J., and Hayes, B. (2003). The impact of IELTS on preparation for academic study in New Zealand. *IELTS International English Language Testing System Research Reports*, 4, 153-206. Canberra: IELTS Australia.



- Shohamy, E. (1993). *The power of test: The impact of language testing on teaching and learning*. National Foreign Language Center Occasional Papers. Washington, DC: National Foreign Language Center.
- Shohamy, E. (2001). *The power of tests: A critical perspective on the uses of language tests*. London: Longman.
- Shohamy, E., Donitsa-Schmidt, S., and Ferman, I. (1996). Test impact revisited: Washback effect over time. *Language Testing*, 13(3), 298-317.
- Spolsky, B. (1999). *Concise Encyclopedia of Educational Linguistics*. Oxford: Pergamon.
- Spratt, M. (2005). Washback and the classroom: the implications for teaching and learning of studies of washback from exams. *Language Teaching Research*, 9(1), 5-29.
- Turner, C. (2001). The need for impact studies of L2 performance testing and rating: Identifying areas of potential consequences at all levels of the testing cycle. In C. Elder, A. Brown, A. K. Hill, N. Iwashita, T. Lumley, T. McNamara, and K. O'Loughlin (Eds.), *Experimenting with uncertainty: Essays in honor of Alan Davies* (pp. 138-149). Cambridge, UK: Cambridge University Press.
- Valette, R.M. (1994). Teaching, testing, and assessment: Conceptualizing the relationship. In C. Hancock (Ed.), *Teaching, testing, and assessment making the connection* (pp. 1-42). Lincolnwood, Illinois: National Textbook Company.
- Wall, D. (1998). Impact and washback in language testing. In Clapham, C. M. and Corson, D. (Eds.) *Language testing and assessment. Encyclopedia of language and education* (pp. 291-302), Vol. 7, Dordrecht: Kluwer Academic Publishers.
- 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. (2005). The impact of a high-stakes testing on classroom teaching: A case study using insights from testing and innovation theory. *Studies in language testing: Vol. 22*. Cambridge, UK: Cambridge University Press/UCLES.

- Wall, D., and Alderson, J.C. (1996). Examining washback: The Sri Lankan impact study. In A. Cumming and R. Berwick (Eds.), *Validation in language testing* (pp.194-221). Great Britain: Cromwell Press.
- Watanabe, Y. (1996). Investigating washback in Japanese EFL classrooms: Problems of methodology. In G. Wigglesworth & C. Elder (Eds.), *The language testing circle: From inception to washback* (pp. 208-239). Melbourne: Applied Linguistics Association of Australia.
- Watanabe, Y. (2004). Methodology in washback studies. In L. Cheng, Y. Watanabe, and A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 19-36). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Wiggins, G. (1994). Toward more authentic assessment of language performances. In Hancock, C. R. (Ed 1994: 69-85) *Teaching, testing and assessment: Making the connection*, Lincolnwood, Illinois USA: National Textbook Company.