

The Washback Effects of High School Third Grade Exam on EFL Teachers' Methodology, Evaluation and Attitude

Mohammad Ahmadi Safa^{1*}, Farhad Jafari²

^{1 & 2} *Bu Ali Sina University, Faculty of Humanities, Department of English Language and Literature, Hamedan, Iran*

Abstract: The widespread use of test scores for different educational and social decision making purposes has made the washback effect of tests a distinct educational phenomenon (Cheng, 1997). The high school third grade final exam in the general educational system of Iran has for long been a high stake test designed to assess the achievement of high school graduates in different school subjects. The present study aimed to investigate the washback effect of this nation-wide exam on EFL teachers' teaching methodology, assessment procedures, and attitudes towards different aspects of the educational system. For this purpose, a researcher made, validated questionnaire was administered to 160 EFL teachers. The results indicated that the third grade nation-wide final exam adversely affects EFL teachers' teaching methodology and increases teaching to the test effect quite noticeably as they try to teach according to the content and format of the test. The results further showed an even stronger negative effect of the exam on EFL teachers' assessment procedures. However, the teachers' attitude towards different aspects of the educational system was not found to be as strongly affected as the other two variables. The findings of the study are of importance for testing and assessment bureaus in charge of extensive high stake tests development. Moreover, raising teachers' awareness of the drawback of teaching-to-the test effect of such a high stake test might help them improve their teaching and evaluation practices.

Keywords: Washback; EFL; Methodology; Assessment; Attitude

Introduction

*Corresponding Author.

Authors' Email Address:

¹ First Author (ahmadisafa@gmail.com), ² Second Author (farhadjafari711@yahoo.com),

ISSN (Online): 2322-5343, ISSN (Print):2252-0198© 2015 University of Isfahan. All rights reserved

Washback or backwash refers to the effect of testing and assessment on teaching and learning processes (Cheng, Watanabe & Curtis, 2004) and follows the idea that tests or examinations can and should drive teaching and learning processes. Interchangeably referred to as measurement-driven instruction, the concept entails a match between the content and the format of the test and the format and content of the instruction (Cheng et al., 2004). The consistency or match has also been termed as curriculum alignment (Shepard, 1990). From another point of view, the test effects and the scope of such effects persuaded Wall (1996) to distinguish between test impact and washback. According to Wall, impact refers to the effects of a test on individuals, policies, or practices in different contexts including the classroom, school, the educational system, and even society at large, while washback/backwash refers to the effects of tests on teaching and learning processes. Washback is inherently believed to move in a particular direction to describe testing–teaching relations; however, Alderson and Wall (1993) identified the bidirectional nature of washback as either positive or negative. Negative washback, the undesirable effect of tests on teaching and learning, happens when there is no match between the assumed goals of teaching and the focus of assessment. On the other hand, positive washback is described as the positive attitude towards the test and cooperative functioning to ensure its assigned purposes. According to Alderson and Wall (1993, p.66) a test has a positive effect “if there is no difference between the curriculum and teaching to test.” From a rather different perspective, Watanabe (2004) described washback in terms of its dimensions, aspects of learning and teaching influenced by the examination, and the factors mediating the process of washback being generated.

Washback effect has attracted great attention in recent years in different educational contexts and has been one of the main lines of research in both general education and foreign or second language educational settings (e.g., Chapman & Snyder, 2000; Cheng, Sun & Ma, 2015; Davies, 1968; Green, 2007; Madaus, 1998; Shih, 2007; Spratt, 2005; Xie, 2015; Zhan & Andrews, 2014).

As an important washback effect of high stake tests and contrary to the perceived common rule proposing that test comes after teaching and learning processes, the priority is inverted in the case of many high stake tests (Cheng, 1997) so that in such testing situations testing comes ahead of teaching and learning. This effect, in turn, influences different aspects and stakeholders of the educational process. As Hughes (1993, p.2) asserts “in order to clarify our thinking on backwash, it is helpful to distinguish between participants, processes, and products in teaching and learning recognizing that all three may be affected by the nature of

the test". Furthermore, that educational systems may be both directly and indirectly affected by high stake tests like school leaving examinations.

Researchers such as Swain (1985) underscored the positive aspects of test effects on language learning and language curriculum. Swain believed that teachers would “*teach to the test*”. In other words, knowing the content and format of the test, the teachers would teach the same or similar content more effectively. Similarly, Wall (2000) believed that the results of the tests’ ‘differentiating rituals’ are, sometimes, so effective in the testees’ future life that the other stakeholders (e.g., teachers) do whatever necessary to help the learners pass the test and the students’ parents ask them to do any possible activities to pass it.

The effects of the tests on teachers and learners are well documented and various studies have examined this effect. It, however, seems that the washback effect of a nationwide high stake test like the third grade school leaving final exam on teachers’ methodology, attitude, and assessment procedures in Iranian high school mainstream educational context is still understudied. As a partial attempt to address the need, this study was conducted to examine the washback effects of the third grade final exam as an annually held exam in the Iranian general education system on specifically EFL teachers’ teaching methodology, assessment procedures, and attitude towards this aspect of the general education.

Review of the related literature

Testing and assessment in versatile forms are integral parts of every system of education. This is why assessment is primarily designed to service teaching and learning (Davies, 1990). However, a role reversal has recently occurred in educational settings because of the impact high stake tests exert on different components of teaching and learning process which has altered teaching to be at the service of testing. This pernicious influence of tests on what goes on in the educational environments and classrooms in particular as well as on the teachers’ teaching procedure has raised some concerns among EFL experts. Additionally, it has given rise to a plethora of studies on the tests and their possible effects on the stakeholders including participants, test developers, and administrators (e.g., Alderson & Wall, 1993; Bailey, 1996, 1999; Chapelle & Douglas, 1993; Damankesh & Babaii, 2015; Hamp-Lyons, 1997; Shohamy, Donitsa-Schmidt & Ferman, 1996; Watanabe, 2004; Xie, 2015).

The aims and scope of washback studies have been quite versatile. Bailey (1996, 1999), for example, proposed that washback should minimally examine both washback to the

program (results of test-derived information provided to teachers, administrators, curriculum developers, counselors, etc.) and washback to learners (the effects of test-derived information provided to test takers) from teachers' and students' perspectives. According to Fulcher and Davison (2007), washback studies should highlight "those things that we do in classroom because of the test, but *'would not otherwise do'* (p.221). Furthermore, in their washback hypothesis, Alderson and Wall (1993, p.117) state that "teachers and learners do things they would not necessarily otherwise do because of the test".

Washback researchers attested that any test may be of both positive and negative effects. According to Wall (2000), positive effect is the drive that persuades testees to cover all subjects completely, complete their assigned syllabuses, and get familiar with other teachers' standards. On the other hand, quoting Wiseman (1961), Wall (2000) maintains that the negative aspect of the test encourages teachers to watch the examiner's foibles and note his idiosyncrasies to prepare students for the most likely test items that might appear in the examination. This negative washback effect restricts teachers' teaching styles and persuades them to concentrate on the 'purely examinable side' of their work and by neglecting other areas. Accordingly, possible positive and negative washback effects of such tests provide ample opportunities and foci for the studies in this realm.

Studies on the washback effects of high stakes tests have shown that these tests make teachers focus on those points that are likely to appear in the tests and teachers usually do not take pedagogical aspects of instructions into account as they usually teach to the test (Hamp-Lyons, 1997). Furthermore, as Bachman (1990) believes, negative washback would result in testing determining the content of teaching. However, it is noteworthy that the extent and nature of test consequences or washback effect depends on teachers' educational background, past learning experience, and beliefs about effective teaching and learning (Watanabe, 2004).

In an effort to further clarify the extent and nature of washback effect, Smith (1991) identified five components of change as a result of washback effects of tests including the target system, the management system, the innovation itself, available resources, and the content in which the change is supposed to happen. Also, Hughes (1993), in his washback model, suggested that participants, processes, and products are the main recipients of the effect. Participants, in Hughes's (1993), included "all of those whose perceptions and attitudes toward their work may be affected by a test". The three elements of the model are described as:

1. **Participants:** Students, classroom teachers, administrators and material developers and publishers whose perceptions and attitudes towards their work may be affected by a test.
2. **Processes:** Any actions taken by the participants which may contribute to the process of learning.
3. **Products:** What is learned and quality of the learning (Hughes, 1993, p.2)

As can be seen in the model, teachers constitute the most noticeable participants in washback studies.

Concerning washback effect type and degree, Alderson and Hamp-Lyons (1996) in their study on the washback effect of TOFEL test reported lots of variations among teachers' perspectives. They maintained that "*our study shows clearly that the TOEFL affects both what and how teachers teach, but the effect is not the same in degree or in kind from teacher to teacher*" (p.295). Contrary to the results Alderson and Hamp-Lyons (1996) obtained, Alderson and Wall (1993) examined the washback effect of innovative tests on Sri Lankan educational system and found that tests can affect content of teaching but less likely they affect the teaching procedure. However, Cheng (1997) in a study on the revised Hong Kong Certificate of Education Examination (HKCEE) found that "*84% of the teachers believed they would change their teaching methodology as a result of the introduction of the revised HKCEE*"(p.45). Similar to Cheng, Lam (1994) supported the washback effect of the tests on the teaching methodology but he further noted that an important factor affecting the methodology change as a result of tests' washback effect is the teaching experience of the teachers. He stated that experienced teachers were much more examination-oriented than their younger counterparts.

Some researchers have investigated the washback effect of different high stake tests on teachers and students' behavior and attitude in the Iranian educational context. Ghorbani (2008), for instance, conducted a survey on the washback effect of University Entrance Examination (UEE) on the teaching practices of a group of pre-university English teachers. Ghorbani examined the six dimensions of classroom activities and time management, teaching methods, teaching materials, syllabus design, teaching content, and classroom assessment. The results showed that all of the participating teachers, regardless of their demographics, were affected negatively by the nationwide high stake UEE.

Contrary to the results reported by Ghorbani (2008), Mousavi and Amiri (2011) investigated the washback effect of Master of Arts level TEFL University Entrance

Examination on the academic behavior of students and professors. They used an observation checklist and two questionnaires to gather the required data. The questionnaires were responded by 32 university teachers and 210 students. They concluded that the test had an insignificant effect on the students and professors' academic behaviors. Nikoopour and AminiFarsani (2012) evaluated the washback effect of State and Azad UEE on Iranian EFL candidates and high school teachers. They found that UEE had influence on teachers' methodology, content of educational programs, students' learning strategies, and teachers' method of evaluation, students and teachers' attitudes and students' affective domain. Furthermore, Razavi Pour, Riazi and Rashidi (2011) investigated the effects of teacher's assessment literacy in moderating the washback effects of summative tests in the EFL context of Iran. For this purpose a test of assessment literacy and a questionnaire on teaching methodology were administered to 53 EFL secondary school teachers. The results revealed that EFL teachers suffer from poor knowledge of assessment and demands of external tests affect their teaching and assessment procedures. Moreover, Nazari and Nikoopour (2011) investigated the washback effects of high school examinations on 120 female Iranian high school learners' language learning beliefs and found that first, the participants agreed on the type of washback effect of the exams and second, there is a correspondence between different factors of learners' language learning beliefs and foreign language learning process.

In another study, Mokhtari and Moradi Abbasabadi (2013) studied the washback effect of Iranian school-leaving tests of English (ISLTE) on teachers' perceptions and performances. They interviewed and observed 10 high school English their classes. The findings verified that ISLTE had a strong negative washback effect on their teaching procedures. The negative washback effect of the test was shown in the form of materials translation by teachers and the absence or disappearance of communicative activities in the observed classes. They suggested that, due to the strong impact of the test on the teachers' teaching methodology, the format of ISLTE was in need of serious revision. Finally, Amengual (2010) examined the washback effects of a high-stakes English Test (ET) on curriculum, materials, teaching methods, and teachers' feeling and attitudes and found that ET clearly affected curriculum and materials.

In addition to University Entrance Examination (UEE) at different levels of BA, MA, and PhD, there are some other nationwide high stake tests held by the Ministry of Education in the context of general education of Iran. One such a test is the third grade high school final examination. The test, as a gate keeping test, plays a determining role in the candidates'

follow up studies. Moreover, the grade point average of the examinations has a direct effect on the high school students' university entrance examination results. Hence, due to the significance of these examinations for both teachers and students, this study was designed to probe into the potential effects of a less frequently studied high stake test on the high school teachers' teaching methodology, testing and assessment procedures, and attitudes towards the educational system. Against this backdrop the following research questions were posed:

RQ1: Does the nation-wide third grade final English examination of high school have any washback effect on English teachers' teaching procedures?

RQ2: Does the final English examination have any washback effect on English teachers' classroom evaluation and assessment procedures?

RQ3: Does the final English examination have any washback effect on teachers' attitude towards different aspects of the educational system?

Method

Participants

One hundred sixty EFL teachers who were teaching third grade courses in high schools were chosen based on convenience sampling procedure to participate in the study. The participating teachers were teaching in the two cities of Malayer and Boroujerd. They were all high school English teachers and either held MA in TEFL (15 %), or BA in English literature, translation or TEFL (85 %). Thirty percent of participants were female and 70 percent were male EFL teachers. Most of the participants had the experience of teaching at different grades or levels of high schools and pre-university centers. Table 1 summarizes the teaching experience and the number of the participants.

Table 1. Participants' Teaching Experience

Teaching experience	N.	Percent
1-5	19	11.75
6-10	30	18.75
11-15	68	42.75
Over 15	43	26.75
Total	160	100

Instruments

The main instrument in this study was a researcher-made five-point Likert scale questionnaire. The first version of the questionnaire was developed based on a few previously designed questionnaires (e.g., Cheng, 1997; Mousavi & Amiri, 2011; Nikoopour & Amini Farsani, 2012) and the ideas the researchers received from some TEFL experts. The first draft included 19 statements to tap the participants' opinion about the three intended areas of the washback effects of the test. The early draft was reviewed by two TEFL experts in order to ensure its content and face validity. The draft was reviewed and revised based on the suggestions and the comments of the TEFL experts.

Afterwards, the questionnaire was piloted with 60 EFL teachers. Analyzing the obtained data through principle component factor analysis (PCA), 6 items (9, 11, 12, 13, 16 and 14) were excluded from the final version of the questionnaire due to poor correlations and factor loadings (less than 0.3). The final version questionnaire included 13 five point Likert scale items ranging from *strongly disagree (with the assumed value of 1)* to *strongly agree (with the assumed value of 5)*. Cronbach's alpha internal consistency reliability index of the questionnaire was estimated to be 0.71 ($\alpha = 0.71$), and hence, deemed acceptable. In addition, Keiser-Mayer Olkin test of adequacy of items was fairly acceptable (KMO= 0.68) and Bartellet's test of Sphercity was significant ($p = .000$). The final 13 item questionnaire was used to tap the participants' ideas on three factors of the EFL teachers' teaching methodology, evaluation procedures and attitudes towards the education system (see Appendix A).

Procedures

The participating teachers in both pilot and main study were met in their schools. Consents were obtained prior to the administration of the questionnaire. At the pilot phase of the study, 60 high school EFL teachers took the questionnaire. The main aim of the piloting stage was to do a validation study on the instrument and estimate the reliability of the questionnaire. After the pilot study and the PCA statistical procedure, a group of 100 high school teachers were asked to take the questionnaire and the collected data were descriptively analyzed to answer the research questions.

Results

Pilot study results

As mentioned before, the first version of the questionnaire was first administered to 60 EFL teachers who were teaching the third grade high school English course. As shown in Table 2, the questionnaire had an appropriate level of adequacy since the observed KMO value exceeded the minimum acceptability level of 0.5 or 0.6, ($KMO=0.68 > 0.5$ or 0.6). In addition, the Bartlett's test of Sphericity was significant showing that the principal component factor analysis was safe to be conducted.

Table 2. KMO and Bartlett's Test Results

Kaiser-Meyer-Olkin Measure of Sampling Adequacy	.682
Bartlett's Test of Sphericity Approx. Chi-Square	675.084
d. f	171
Sig	.000

As in Tables 3 and Table 4, the factor analyses confirmed the strong correlation of the questionnaire items with three main factors.

Table 3. Principle Component Analysis Results of the Questionnaire

Component	Initial Eigenvalues			Sums of Squared Loadings		
	Total	% of variance	Cumulative %	Total	% of variance	Cumulative%
1	3.926	30.197	30.197	3.926	30.197	30.197
2	2.467	18.980	49.177	2.467	18.980	49.177
3	1.539	11.842	61.019	1.539	11.842	61.019
4	.961	7.391	68.409			
5	.897	6.901	75.310			
6	.708	5.445	80.755			
7	.610	4.691	85.446			
8	.540	4.153	89.599			
9	.328	2.520	92.120			
10	.318	2.448	94.568			
11	.281	2.158	96.726			
12	.221	1.697	98.424			
13	.205	1.576	100.000			

In addition, the initial eigenvalues of only the first three components exceeded the criterion value of 1 (Pallant, 2013) and the cumulative percentage of the three components explained a total of 60.01 percent of the variance.

As seen in Table 4, the variance was divided among 13 items and 6 items out of the total of 19 questionnaire items were discarded due to poor correlation and factor loading (less than 0.3). Finally, the rotated factor matrix identified the more strongly correlated items for each factor. As such, the final questionnaire including 13 variables which tapped altogether the three main factors was achieved. The three factors were named as *methodology (factor 1)*, *attitude (factor 2)* and *evaluation (factor 3)*. Items 8, 3, 7, 2, and 5 were loaded on factor one, items 15, 17, 18, 19 tapped factor two and factor three was tapped by items 1, 10, 4, and 6 of the 19 item questionnaire. Meanwhile, it should be noted that as some of the questionnaire items (items 2, 15, and 14) were negatively correlated with the factors, hence, reversely computed and analyzed.

Table 4. Rotated Factor Matrix

Items	Factor		
	1	2	3
6) Using final exam items in mid-term exams	.108	-.038	.908
5) Explaining about the content or type of final exam items.	.689	-.144	.069
2) In private institutes, I would use the same methods and techniques I am using now.	-.656	-.023	-.214
7) I tend to teach my students so that they can pass	.562	-.038	.149
3) Teaching the material according to their importance.	.556	-.047	.494
8) Teaching students the tips and tricks to answer the final exam.	.474	-.055	.132
13) The exam helps students' thinking style and creativity.	.183	-.074	-.038
16) I tend to teach and use tactics for answering multiple choice items	.162	-.102	-.79
18) I am satisfied with language testing procedure in Iran.	-.011	.887	-.087
19) I think I must prepare my students to have an active role in society.	.022	.877	-.077
17) My students themselves should choose their field of study.	-.100	.755	-.024
15) I think the final exam causes fear and stress.	.175	-.439	-.006
12) The exam reflects the educational aims for the courses.	-.191	.220	-.009
4) My testing method helps students to get ready for exam.	.066	.001	.679
1) I consider third grade final exam, while testing and teaching.	.269	-.234	.671
10) This exam has positive effect on the whole education process.	.230	-.177	.445
9) I practice only those points similar to the final exam.	.082	.042	-.294
11) As the exam is an effective test, I prepare my students for it.	.075	.071	.257
14) I attract students' attention to classified teaching materials.	-.089	.026	.162

Main Study Results

In the second phase of the study 100 EFL teachers took the validated 13 item questionnaire. The data collected were analyzed in terms of descriptive analyses and frequency counts. The frequency for each level of the Likert scale of the items of each factor was computed and the average frequency for each was obtained. In addition, in order to obtain the mean value of the responses to each questionnaire item, considering the assumed value of the levels of the scale (strongly disagree =1, disagree=2, undecided=3, agree=4, and strongly agree=5), the mean value for all items was calculated and finally the average mean score for the factor was obtained.

As Table 5 presents, about 72 percent (45.8+26.2) of the teachers believed that the exam affected their teaching methods in EFL classes and the average mean score for this factor was fairly high (3.84).

It is necessary to add that the questionnaire items in the following tables were minimally presented due to the limited space of the tables and the full report of the questionnaire items is presented in the appendix.

Table 5. The Washback Effect of the Exam on Teachers' Methodology (factor 1)

Items	St. Disagree	Disagree	Undecided	Agree	St. Agree	Mean
2- In private institutes, I would use the same methods I am using now.	11	5	4	43	37	3.9
3- I teach the material according to their importance	1	6	18	42	33	4
5- In my class, I explain about the content final exam items.	1	9	18	53	19	3.8
7- I tend to teach my students so that they can pass the final exam.	0	7	23	49	21	3.8
8- I teach the students the tips and tricks to answer the final exam.	2	13	22	42	21	3.6
Average	3	8	17	45.8	26.2	3.84

Evidently, 43 percent of the respondents agreed that if they were to teach a third grade final exam preparatory course (item 2) , they would use the same methods and techniques

they were using in their regular classes of high schools in which the academic skills and abilities are to be given the first priority. Thirty seven percent “strongly agreed” with item 2 which means that, added to the percentage of the teachers who “agreed” with this statement, 80 percent of the teachers “change their teaching methodology so that they could prepare their learners for the test in an attempt to guarantee their learners’ success at the intended test”. Items 7 and 8 of the questionnaire referred to the teaching tips and tricks for successful test taking of the learners and ultimate success in passing the test. The percentage of responses to these questionnaire items at different points of Likert scale were quite revealing (Table 5) confirming the existence of teaching to the test process in the studied educational context. On the other hand, 42 percent of the respondents ‘agreed’, and 33 percent ‘strongly agreed’ with the statement in item 3 of the questionnaire stating that “I teach the material and learning points according to their importance level in the exam”. This means that the content of teaching was also strongly affected by the test content as well, as altogether 75 percent of the respondents accepted the stated rationale for the choice of the content of their teaching.

Table 6 presents the descriptive statistic information for the ‘evaluation’ factor. A total of about 84 percent of the teachers either agreed (49 %) or strongly agreed (35 %) that the test exerted a significant effect on their evaluation and assessment procedures. The total mean score for this factor (4.12) compared with the methodology factor (3.84) appeared to be significantly higher which was indicative of even stronger influence of the test on the teachers’ evaluation and assessment procedures.

Table 6. The Washback Effect of the Exam on Teachers’ Evaluation Procedures (factor 3)

Items	St. Disagree	Disagree	Undecided	Agree	St. Agree	Mean
1- I consider third grade final exam while testing and teaching in my classes.	1	5	2	43	49	4.34
4- I think my testing methods are helping students to get ready for exam	0	3	9	53	35	4.2
10- This exam has positive effect on whole education process.	0	7	14	46	33	4.05
6- I use final exam items in my mid-term exams.	0	8	15	54	23	3.9
Average	0	6	10	49	35	4.12

As is evident in Table 6, a total of about 92 percent of the teachers either agreed (43 %) or strongly agreed (49 %) with the first questionnaire item (item 1) saying that *I consider third grade final exam while teaching and testing in my classes*. A mean score of 4.34 was

clearly indicative of the strength of the effect of the test on the addressed areas of the teaching and assessment. More or less similar effect was evident above for the other items. Roughly speaking, items 1 and 4 considered *how* of testing and 6 and 10 focus on *what* of testing. In other words, the two categories of items addressed the content and the procedure of testing the teachers use in their educational context. The percentages and the mean scores presented in Table 6 are strongly indicative of the influence of the third grade final exam on both what and how of the teachers' assessment and testing.

Finally concerning the test's washback effect on the teachers' attitude, the obtained results, presented in Table 7, show that totally about 45 percent of participants believed that the test affected their attitudes significantly (29% agreed, 16% strongly agreed); however, 19 percent were undecided, and 36 percent denied the tests' impact (28 % disagree + 8 % strongly disagree) in this regard. The average mean score for this factor (3.16) was the lowest compared to the other two factors i.e., teaching methodology and evaluation procedures.

Table 7. The Washback Effect of the Exam on Teachers' Attitude (factor 2)

Items	St. Disagree	Disagree	Undecided	Agree	St. Agree	Mean
15- I think the final exam causes fear.	8	23	30	26	13	3.13
17- My students themselves should choose their field of study.	2	34	16	42	6	3.16
18- I am satisfied with language testing procedure in Iran.	10	31	14	27	18	3.12
19- I think I must prepare my students to have an active role in society.	12	23	18	21	26	3.26
Average	8	28	19	29	16	3.16

The comparisons for the descriptive statistic information gained for the three factors are presented in Table 8. According to the obtained results, the teachers' evaluation and assessment procedures were highly affected by the nation-wide third grade high school final exam and the magnitude of the test's washback effect on the teachers teaching methodology was placed in the second place of importance. However, it seems that the participants' attitude towards the Iranian general education system was not highly affected by the test as the total mean score obtained for this factor (3.16) was fairly close to the mid position of the Likert scale that was neutral in value. Consequently, the findings roughly indicated that the

first and second null hypotheses of the study which denied any kind of washback effect of the test on the teachers' teaching methodology and evaluation and assessment procedures were both rejected while the third hypothesis which rejected the effect of the test on the teachers' attitude was confirmed.

Table 8. The Washback Effect of the Exam on EFL Teachers' Methodology, Attitude, and Evaluation

Factors	St. Disagree	Disagree	Undecided	Agree	St. Agree	Mean
1- Teachers' methodology	3	7	16	47	27	3.84
2- Teachers' attitude	8	28	19	29	16	3.16
3- Teachers' evaluation procedures	0	6	10	49	35	4.12
Average	4	14	15	41	26	3.70

Discussion

Testing and assessment as integral parts of education play a wide range of prognostic and diagnostic roles in education process and help the pedagogical or educational processes which might precede or follow them. However, their positive contribution to education is not free of some negative effects on the same processes. Both positive and negative effects of testing on the follow up teaching and learning processes have been termed as washback or backwash by the testing and assessment scholars (e.g., Alderson & Wall, 1993; Hughes, 1993; Wall, 1996). However, the present concern with washback was ignited by Messick's (1989) introduction of the notion of consequences into his definition of validity (Fulcher, 2010). While the existence of washback effect is not in question, the *how* of this effect is not so clear (Tzagari, 2009) and hence needs to be studied. The need for the study of the washback effects of high stake tests is clearly more significant than the same need for low stake tests due to the wider scope of the consequences accompanying such tests. High stake tests are considered and used as agents of change (Luxia, 2005); however, as many empirical studies have shown and the stated results of the present study confirmed, the use of high stakes tests is not usually as effective as they are planned (Qi, 2004, as cited in Fulcher, 2010) and sometimes not in the same way as their designers meant (Andrews, 1994). The current study partially attempted to address the washback effects of a high stake nationwide achievement test that is administered by the end of third year of high school in the Iranian general education system.

By the end of the third year of secondary high school all subject matters taught during the educational year are subject to this nationwide evaluation through which students across the country take a single test for each subject at exactly the same time. The test results are influential in the candidates follow up academic studies in higher education centers, colleges, and universities. The washback effects of the high stake test of English as a Foreign Language (EFL) on the English teachers' teaching methodology, assessment and evaluation methods, and attitudes towards educational processes were studied in this piece of research. The washback effect of the test on the teacher variables was focused on here as teachers are highly decisive and hence most visible participants in washback studies among other participants (Baily, 1999) owing to the direct effect of tests on their pedagogical behaviors.

In this study, a researcher-made and validated questionnaire was administered to 160 EFL teachers who were teaching English courses of the third grade of high schools in the pilot and main study phases. The results indicated that the third grade nationwide final test significantly affects EFL teachers' teaching methodology and increases teaching to the test effect. As is described above, the participating teachers' teaching methods were be under the negative impact of the test since they stated that they change their teaching method so that they could guarantee their students' success at the test. This point completely confirms Alderson and Wall (1993, p.117) who stated that "*teachers do things they would not necessarily otherwise do because of the test*". The participants also openly agreed with the focus on the teaching *tips and tricks of taking the test* to the sacrifice of the academic and pedagogical aspects and content of the course. Hamp-Lyons (1997) referred to a similar point when he noted that these tests made teachers focus on points that were likely to appear in the tests and they usually did not take into account *pedagogical aspects of instruction* which meant that they taught to the test. This finding is also consistent with that of many previous studies such as Alderson and Hamp-Lyons (1996), Watanabe (1996), Cheng (1997), Luxia (2005), Spratt (2005), Ghorbani (2008), Nikoopour and AminiFarsani (2012), SeyedErfani (2012), Zhan and Andrews (2014) and Damankesh and Babaii (2015) all of which confirmed the negative impact of high stake tests on the teaching methodology of teachers.

However, the finding confirming the negative impact of high stake tests on teaching methodology was not consistent with few studies such as Alderson and Wall (1993) who concluded that the tests influenced the content of teaching but had no impact upon teaching methodology. Similarly, Shin (2009) suggested that teachers' instruction was not vulnerable to the test impacts it exerted the micro-level contextual factors and teacher factors.

In addition to the *how* of teaching which was affected by the third grade nationwide test, the content or *what* of teaching of the participating teachers was also highly affected, as the majority of the teachers (75 %) chose their teaching practice based on the test content. Conversely, Wall (2000) maintained that one of the negative washback effects of the tests happened when the teacher prepared the test takers for the most likely test items that might appear in the examination. This negative washback effect would persuade them to concentrate on the ‘purely examinable side’ of their work and the other areas to be overlooked. The lack of attention to the other pedagogical aspects excludes the possibility of *measurement driven instruction* (Cheng & Watanabi, 2004) which favors a match between the content and format of the test and the format and the content of the instruction. In measurement driven instruction the regular course of instruction is to be reflected in the format and content of the test while teaching to the test entails a ‘role reversal’ (Davies, 1990) in that it is the content and format of the test that controls the process and content of the preceding instruction. In other words, teaching is at the service of testing (Cheng, 1997; Davies, 1990) while it is believed to be the other way round. The reported negative washback effect of the high stake tests on the content or what of teaching confirms the earlier studies results (e.g., Alderson & Hamp-Lyons, 1996; Hamp-Lyons, 1997; Ghorbani, 2008; Nikoopour & AminiFarsani, 2012; and Cheng, Sun & Ma, 2015).

Furthermore, the results of the present study confirm an even stronger significant negative effect for the high stake test on the EFL teachers’ testing and assessment procedures since an absolute majority of the respondents (92%) verified that they consider both the format and the content of the high stake tests in their own testing and evaluation practices. It is concluded that both what of testing and how of testing are affected by the high stake test. The effect on the teachers’ assessment and testing procedures seems to be even stronger than the effect on the teachers’ methodology. This point further supports the finding that the teachers do whatever that familiarizes the learners with the content and format of the high stake test and prepares them for it while they might not embark on the same course of teaching, testing and other pedagogical activities if it was not for the sake of the test or if the test did not exist (Alderson & Wall, 1993). In other words, not only teaching to the test is practiced but also ‘testing to the test’ is quite evident. Other already referred to researchers like Hamp-Lyons (1997), Ghorbani (2008) and Nikoopour and AminiFarsani (2012) have also reported the negative impact of high stake tests on the testing, assessment, and evaluation procedures of the teachers.

Finally, the last finding of this study verifies that, unlike teachers' teaching and testing methodology, their attitude towards different aspects of general education including teaching and learning processes are as strongly affected by the test as the other two factors. This, in turn, indicates that the teachers were applying quite strategic pedagogical practices to achieve the most practical and institutionally valued objective that is to enable their learners pass the test, while their attitude towards the desirable educational processes are not deeply affected and altered. A probable explanation for this effect might be the fact that the tests' differentiating rituals (Wall, 2000) sometimes are so effective in the testees' future life that the teachers ask the testees to do any possible activities to *only* pass the tests and quite clearly they change their own pedagogical practices to serve this purpose. This last finding seems to be in complete accord with the previous findings as the existence of negative washback effect projects the lack of positive attitude of the stakeholders towards the test. Alderson and Wall (1993) believed that a positive washback would function when there is a positive attitude toward the test and there is a cooperative working to fulfill its assigned purposes. The negative washback effect of the third grade nationwide English exam on the teaching and assessment procedures of the teachers is indicative of the lack of a positive attitude of the teachers towards the test.

Conclusion

The results of this specific study provide evidence to the fact that content and format of teaching are to a great extent geared towards and adapted to high take tests content and format. Both *what* of teaching and *how* of teaching of the EFL teachers were negatively affected by the content and format of the specific studied high stake test. It verifies the results of previous studies on the washback effects of high stakes tests on teachers attitude and methodology in EFL classes. However, it is emphasized that EFL teachers spend most of their class time to practice the material which are likely to be included in third grade final exam and the communicative skills of the language which were not likely to be included in the studied exam were all neglected. Evidently, this procedure has negative and detrimental effects on the overall foreign language communicative competence development of Iranian high school students as the EFL teachers did not prioritize this main aspect of foreign language learning over the language related components which were deemed to be included in the third grade final exam. In addition, EFL teachers' classroom assessment procedures and evaluation format were so designed that maximum similarity with the content and format

of the high stake tests was achieved, maximally preparing them for the test in advance. All such pseudo-pedagogical activities were carried out as the teachers were committed to do everything to help their learners pass the test successfully, even though the true learning did not take place and the academic and educational goals were not achieved. In other words, teaching served testing and the teachers did “*teach to the test*” despite the fact that they were aware of the harmful effects of this behavior. To counteract these potential negative washback effects, as Shohamy (1993, p.187) argues, a continuous and cooperative loop between external test developers and people working in the schools seems to be vital. The study results necessitate a number of changes in the program. The changes include not only changes in the test content and format and the testing procedure but also, as Lam (1994) rightly attests to, changes in the teaching culture.

References

- Alderson, J.C., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14, 115-129.
- Alderson, J. C., & Hamp-Lyons, L. (1996). TOEFL preparation courses: A study of washback. *Language Testing*, 13, 280-297.
- Amengual, M. (2010). Exploring the washback effects of a high stakes English Test on the teaching of English in Spanish upper secondary schools. *Revista Alicantina de Estudios Ingleses*, 49, 149-170.
- Andrews, S. (1994). The Washback effect of examinations: Its impact upon curriculum innovations in English language teaching. *Curriculum Forum*, (1), 44-58.
- Bachman, L.F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bailey, K.M (1996). *Washback in language testing*. Princeton, NJ: ETS.
- Bailey, K.M (1999). Working for washback: A review of the washback concept in language testing. *Language Testing*, 13(3), 257-279.
- Chapelle, C., & Douglas. D. (1993). Foundations and directions in new decade of language testing. In D. Douglas & C. Chapelle (Eds.), *A new decade of language testing research* (pp.1-22). Alexandria, VA: TESOL Publications.
- Chapman, D.W., & Snyder, C.W. (2000). Can high stake national testing improve 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 in Education, 11*(1), 38-54.
- Cheng, L., Watanabe, Y., & Curtis, A. (Eds.) (2004). *Washback in language testing: Research context*. Mahwah, N. J: Lawrence Erlbaum.
- Cheng, L., Sun, Y., & Ma, J. (2015). Review of washback research literature within Kane's argument-based validation framework. *Lang-Teach, 48*(4), 436-470.
- Damankesh, M., & Babaii, E. (2015). The washback effect of Iranian high school final examinations on students' test-taking and test-preparation strategies. *Studies in educational evaluation 45*, 62-69.
- Davies, A. (1990). *Principles of language testing*. Oxford: Blackwell.
- Davies, A., (1968). *Language testing symposium: A psycholinguistic approach*. Oxford: Oxford University Press.
- Fulcher, G. (2010). *Practical language testing*. London: Hodder Education.
- Fulcher, G., & Davison, F. (2007). *Language testing and assessment: An advanced resource book*. New York: Routledge.
- Ghorbani, M. R. (2008). *Washback effect of university entrance examination on Iranian pre-university English language teachers' curriculum planning and instruction*. (Unpublished doctoral dissertation, University Putra Malaysia). Retrieved from <http://psasir.upm.my/4901/>.
- Green, A. (2007). Washback to learning outcomes: A comparative study of IELTS preparation and university pre-session language courses. *Assessment in Education, 14* (1), 75-97.
- Hamp-Lyons, L (1997). Washback, impact and validity: Ethical concerns. *Language Testing, 14*(3), 294-303.
- Hughes, A. (1993). *Backwash and TOFEL 2000*. (Unpublished manuscript). University of Reading.
- Lam, H.P. (1994). Methodology washback—an insider's view. In D. Nunan, R. Berry, & V. Berry (Eds.), *Bringing about change in language education: Proceedings of the International Language in Education Conference 1994* (pp.83-102). Hong Kong: University of Hong Kong.
- Luxia, Q. (2005). Stakeholders' conflicting aims undermine the washback function of high stake test. *Language Testing, 22*(2), 142-173.

- Madaus, G.F. (1998). The influence of testing on curriculum. In Tanner, L.N (Ed.), *Critical issues in curriculum* (pp.83-121), Illinois: University of Chicago Press.
- Mokhtari, S. A. & MoradiAbbasabadi, M. (2013). Examining the influence of the Iranian school leaving test of English (ISLTE) on teachers' perceptions and performances. *International Journal of Linguistics*, 5 (2), 1-23.
- Mousavi, M., & Amiri, M. (2011). The washback effect of TEFL university exam on academic behavior of students and professors. *Journal of English Studies*, 1 (2), 103- 144.
- Nazari, M. & Nikoopour, J. (2011). The washback effect of high school examinations on Iranian learners' language learning beliefs. *Journal of Language and Translation* 2 (1), 29-49.
- Nikoopour, J. & AminiFarsani, M. (2012). Depicting washback in Iranian high school classrooms: A descriptive study of EFL teachers' instructional behavior as they relate to university exam. *The Iranian EFL journal*, 8(1), 9-34.
- Pallant, J. (2013). *SPSS survival manual: A step by step guide to data analysis using IBM SPSS (5th ed.)*. New York : The Mc Graw Hill Companies.
- Razavi Pour, K., Riazi, A., & Rashidi, N. (2011). On the interaction of test washback and teachers assessment literacy: The case of Iranian EFL secondary school teachers. *English Language Teaching* 4(1), 156-161.
- SeyedErfani, S. (2012). A comparative washback study of IELTS and TOEFL IBT on teaching and learning activities in preparation courses in the Iranian context. *English Language Teaching*, 5(8), 185-195.
- Shepard, L. (1990). Inflated test score gains: Is it old norms or teaching to the test. *CSE Technical Report 307*. Los Angeles, CA: University of California at Los Angeles.
- Shih, C. (2007). A new washback model of students' learning. *The Canadian Modern Language Review*, 64 (1), 135-162.
- Shin, C-M. (2009). How tests change teaching: A model for reference. *English Teaching: Practice and Critique*, 8 (2), 88-206.
- Shohamy, E. (1993). A collaborative/diagnostic feedback model for testing foreign languages. In D. Douglas & C. Chapelle (Eds.), *A new decade of language testing research* (pp. 185-202). Alexandria, VA: TESOL Publications.
- Shohamy, E., Donitsa-Schmidt, S., & Ferman, I. (1996). Test impact revisited: Washback effect over time. *Language Testing*, 13 (3), 298-317.
- Smith, M.L. (1991). Put to the test: The effects of external testing on teachers. *Educational Researcher*, 20 (5), 8-11.

- 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.
- Swain, M. (1985). Large scale communicative language testing: A case study. In Y.P. Lee, A. C. Y. Y. Fok , R. Lord, & G. Low (Eds.), *New direction in language testing*(pp. 35-46). Oxford: Pergamon.
- Tsagari, D. (2009). *The complexity of test washback: An empirical study*. Frankfurt: Peter Lang.
- Wall, D. (1996). Introducing new tests in to traditional systems: Insights from general education and from innovation theory. *Language Testing*, 13(3), 334-354.
- Wall, D. (2000).The impact of high-stakes testing on testing and learning: Can this be predicted or controlled? *System*, 28, 499-509.
- Watanabe, Y. (1996). Does grammar translation come from the entrance examination? Preliminary findings from classroom based research. *Language Testing*, 13, 318-333.
- Xie, Q. (2015). Do component weighting and test method affect time management and approaches to test preparation? A study on washback mechanism. *System*, 50, 56-68.
- Zhan,Y & Andrews, S. (2014). Washback effects from a high-stakes examination on out-of-class English learning: Insights from possible self-theories. *Assessment in Education: Principles, policy & practice*, 21(1), 71-89.

Appendix

Questionnaire: The researchers highly appreciate the time you spend to take the questionnaire.

Age:.....**Gender:** *Male* *Female* **Teaching Experience:**years

Education Level: BAMAPhD

Strongly disagree (1), disagree (2), undecided (3), agree (4), strongly agree (5)

	Statements	1	2	3	4	5
1	I consider third grade final exam, while testing and teaching in my classes.					
2	If I were supposed to teach in a final exam preparation course in private institutes, I would use the same methods and techniques I am using now.					
3	I teach the material and learning points according to their importance in final exam.					

- 4 I think my testing methods are helping students to get ready for final exam.
 - 5 In my class, I explain about the content or type of final exam items.
 - 6 I use final exam items in my mid-term exams.
 - 7 I tend to teach my students so that they can pass the final exam.
 - 8 I teach the students the tips and tricks to answer the final exam.
 - 9 I tend to practice only those points that are similar to the points tested the final exam.
 - 10 This exam has positive effect on the whole education process.
 - 11 As the exam is a quite effective test, I prepare my students for it.
 - 12 The exam reflects the educational aims and purposes set for the courses.
 - 13 I think the exam helps to enhance my students' thinking style and creativity.
-