ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202205_5(5).0010

The Washback Effect of Listening Parts of TEM on Teaching Affairs and Test Marking

Xiaoqin Shi[1, a], Xiaoyuan Wang[1, b], Renyi Deng[1, c]

[1]The Department of Foreign Languages, Northwest A&F University, Yangling, Shaanxi Province, 712100, China

[a]sxq-shi@nwsuaf.edu.cn, [b]924528875@qq.com, [c]6057978@qq.com

Abstract

TEM is a criterion-referenced standardized test for English major students of colleges in China and it examines the results of teaching and learning. The score of TEM becomes a scale of teaching level and student English capability, which is admitted by society. The washback effect of TEM on teaching performance caught most language researchers' eyes during a few decades. However, few empirical studies conducted a vertical comparison of the score of TEM test to illustrate the washback effect on teaching affairs and test making. The study aims to explain how the washback affects teaching affairs, like training program and curricula arrangement in NWAFU, and test making by comparing student TEM listening scores of 9 years from 2010 to 2019. It includes vertical comparisons of total 9-year listening scores of three groups, such as Science and Technology College (STC), National College (NC), and Northwest A&F University (NWAFU). The results are the trend of listening score rises up during nine years for all groups, and the washback impacts on NWAFU teaching affairs and test making strongly and positively. The study will give teaching affair office and test committee some suggestions about taking advantage of the washback of TEM to arrange listening courses and test form. The further study will explore how the washback effect of other TEM parts impacts on the student learning and one empirical study will be conducted.

Keywords

Washback; Listening parts; NWAFU.

1. Introduction

Test for English Major (TEM), conducted by the National Advisory Committee for Foreign Language Teaching, Ministry of Education of China, has two stages, TEM 4 for sophomores and TEM 8 for seniors of college English major. It examines whether students can meet the requirement of Teaching Syllabus and most scholars think it is also a measure of English teaching quality. Lots of students long for pass the TEM to show their English proficiency. Teachers try to exert their abilities to teach and improve student's pass rate of TEM. TEM has become the most important test for English major students and a key examine of teaching results. So, many language researchers study the relationship between the test and teaching and learning. And the washback effect of test on teaching and learning is one of direction of their studies.

Washback is also called backwash (Hughes, 1989) which refers to the influence of testing on teaching and learning (Alderson & Wall, 1993; Bailey, 1996; Buck, 1988; Shohamy, 1996). Scholars defined washback as a) positive washback and negative washback. Hughes (1989) thought positive washback can promote teaching and learning, while negative washback means a test can hamper teaching and learning. For example, students and teachers may take more attention to test papers or test skills, and get rid of textbooks and syllabus; b) Prodomou (1995)

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202205_5(5).0010

depicted washback as covert or overt washback. He said overt washback is negative, like aforesaid example, while covert washback, implicit effect of test on teaching and learning, makes teaching activities like a test procedure. There are plenty of studies about washback, and famous theoretical washback are Alderson & Wall's Washback Hypothesis, Hughes' PPP Model, Shih's Washback Model and Watanabe's dimensions of washback. Alderson & Wall (1993) put forward 15 hypothesis and provided a solid foundation for further empirical study on washback. Hughes (1993) proposed trichotomy Model, that is to say PPP (participants, processes, and products) model. Shih (2007) examined the GEPT washback effect on student's learning and concluded washback effect was from three aspects: external factors, internal factors, and test factors. Watanabe (1992) subdivided washback into five dimensions: 1) Specificity. Washback can be general or specific; 2) Intensity. Washback can be strong or weak; 3) Length. The effect of washback can be long-term or short-term; 4) Intentionality. Intended washback means the effects of tests which test designers have presumed on teaching and learning. Whereas unintended washback is the effects beyond presumption; 5) value. There may be positive or negative washback. Andrews, Fullilove & Wong (2002) investigated whether the test of UE promoted students' English proficiency and the study showed that the influence of UE on students' oral performance varied from students to students. Erfani (2012) carried out a study to compare the different impacts of IELTS and TOEFL on teaching and learning and his finding TOEFL's impact covered most teaching and learning activities, but IELTS impacted on oral test more deeply.

Besides, Chinese scholars did many of washback studies on some high-stake exams or national examinations. Li (1997) carried out washback investigation from NMET (National Matriculation English Test in China). The study showed that washback depends on feedback, and the power of test is to inform and influence. Qi (2011) also conducted her study of NMET washback. She summarized that NMET improves student's English comprehensive ability, but teachers should give more oral practices in class. Huang (2002) tried to find out the washback of CET-4 (College English test band 4) on college English teaching and learning. He said that CET exerted influences on multiple aspects of learning and teaching, and CET's washback was more positive than previously presumed. Gu (2007) conducted a longitudinal study of CET washback, covering a huge number of aspects of CET washback and it is called an encyclopedic guide book for further CET washback study. There are also lots of CET washback study on CET special parts, such as Cao (2009) delved into listening part, Yu (2012) focused on the part of writing and Dong (2009) dealt with reading.

Although TEM is not as popular as CET and NMET in China, scholars also pay more attentions to its washback study. He (1998) studied TEM washback on English grammar teaching and he concluded that grammar test has a positive effect on teaching. However, he did not give an empirical illustration. The following studies of TEM washback are most from a macro-perspective and scholars talked more about TEM washback positive or negative effect on teaching (Liu 2005; Xu 2014; Yan, 2011; Kong, 2014), Huang (2004) studied oral test of TEM4, which is from a micro-perspective. All above TEM washback studies took placed before 2016. After 2016, TEM had a big change in test form and content. So, reformed TEM catches scholars' eyes (Zhang, 2017; Wu, 2019;) and they start to focus on how reformed TEM impacts on teaching and learning, especially on students' TEM scores (Hong & Yi, 2016; Wang & Xia, 2018, Zhang, et. al, 2019). Moreover, more and more studies are from a certain part of TEM, like Liu & Hu (2018) studied reading section of TEM and explained how to teach students getting high score in the test, and Meng & Li (2019) conducted an in-depth analysis of several authentic TEM4 grammar and vocabulary test items in the past 15 years. They disclosed the overall tendency on the features of TEM4 test items and gave some suggestions on test preparation and teaching. Although scholars did much work on TEM washback study in China, few studies have been done to illustrate what the trends of TEM scores are in recent years and how the

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202205_5(5).0010

washback of TEM impacts on teaching affairs and test making through vertical comparison of students TEM results and changes of test form and content in recent years. This study will fill the gap in this aspect.

2. Methods

2.1. Research Question

According to above-mentioned literature review, TEM can influence plenty of aspects of English teaching and learning. Based on the theoretical background of washback effect, the study explores the washback effects of listening parts (dictation part and listening part, two parts of six parts of TEM) on teaching affairs and test making. In the study, three groups are three kinds of colleges in China, including Science and Technology College (STC), National College (NC), which includes all colleges and universities of China, and Northwest A&F university (NWAFU), a comprehensive university in China. The explored questions are following:

(1) What is the trend of scores of TEM in nine years, according to average scores of students from three groups (STC, NC, and NWAFU) respectively?

(2) How will the washback of listening parts of TEM be on teaching affairs? Strong or weak?

(3) How will the washback of TEM listening parts impact on test making? Positive or negative?

2.2. The History of TEM

Table 1. The difference of TEM form in the second and third stage

TEM band	parts	Year span	Question count	Part name	type	number	points	Weight(%)	Time (min)
four	I	2010-2015		Dictation	selected-response	1	15	15%	15
		2016-2019				1	10	10%	10
	II	2010-2015	1—30	Listening A. conversation	selected-response	10	15	15%	15
				B. essay		10			
				C. news		10			
		2016-2019	1—20	Listening A. lectures	constructed -response	10	20	20%	20
				B. conversation	selected-response	10			
	III	2010-2015	31-50	Close		20	10	10%	15
		2016-2019	21—40	Grammar & vocabulary		20	20	20%	10
	IV	2010-2015	51-80	Grammar & Vocabulary	selected-response	30	15	15%	15
		2016-2019	41-50	close		10	10	10%	10
	V	2010-2015	80-100	Reading		20	20	20%	25
		2016-2019	51—60	reading A.	selected-response	10	10	20%	20
			61—65	reading B.	constructed-response	5	10		
	VI	2010-2015		Writing A. composition	constructed-response	1	15	15%	35
				B. note writing		1	10	10%	10
		2016-2019		Composition		1	20	20%	45
total		2010-2015				103	100	100%	130
		2016-2019				67	100	100%	130

Test for English Major (TEM) started in 1990. During the 20 years, the test underwent three stages. The first stage was from 1990 to 2003. In 1997, the syllabus (revised edition) of TEM was published officially and the test content was composed by 90 questions and included six parts: writing, dictation, listening, cloze, grammar and vocabulary, and reading. The limited time of TEM was 140 minutes. The second stage was from 2004 to 2015. Although the test form changed, it was not a big. The number of test question was added to 100 and the order was reversed. Part I became the part VI, the last part of TEM and the limited time was 130 minutes. However, the syllabus of TEM has been reformed in the third stage and TEM form and content had a big change after 2015, especially in part I and part II, dictation and listening. Table 1 shows the structure of TEM in the second and third stages.

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202205_5(5).0010

2.2.1. Changes of Listening Parts

The study focuses on listening parts (part I and part II). Some changes are shown in Table 1.

Change in part I (Dictation parts): In dictation part, the points and weight are reduced from 15, and 15% to 10, and 10%, as well as limited time from 15 to 10 minutes. The length of dictated article also changes, from 150 words to 80-90 words. The subject of the article is broader than that in the second stage.

Change in part II (listening part): This part has more changes, compared with that in second stage. First, the question form is adjusted from single-choice questions to the combination of answer and multiple-choice ones. Second, the content of this part changed from "conversation", "essay" and "news" to "lectures" and "conversations". In section A, Students need to fill in the blank to answer the questions after they listen to mini-lecture which contains about 150 words. The test time is 10 minutes and the points and weight are 10 and 10%. Section B of this part is "Conversation" consisted of two conversations about 450 words. There are 10 multiple-choice questions in section B. Students are asked to select a best answer from the four choices given in the test paper. Test time, score and weigh are the same as section A. It seems that the listening part is more difficult in the third stage than that in second stage and students would have gotten lower score, compared with former listening part form. The study will explore if it is true.

2.2.2. The Importance of TEM

Because TEM is a nationwide examination in China, and it is so important for English major students that sophomores from all colleges across country don't want to miss it every year. However, every English major student only has two chances to participate in TEM which takes place in April before the oral test in May. Northwest A & F University (NWAFU) is a "Double Top" (World-class university and first-class disciplines) university in China. Every NWAFU's English major sophomore takes TEM in his second college year. TEM scores of NWAFU students will be a miniature of the average level of comprehensive colleges in China. To explore the trend of the ten-year scores of TEM in three groups will explain how the washback of TEM impact on teaching affairs and test making dynamically.

2.3. Data Collection and Analysis

2.3.1. Data Collection

Although TEM has six parts, the study only refers to student's listening ability, including dictation and listening parts of TEM. The data was collection from teaching affair office of NWAFU, composed of nine years scores of TEM from three groups, STCs, NCs and NWAUF respectively. And students taking TEM are all sophomores of colleges, while juniors' score are not in the study. Because the TEM score sheet of 2011 is missing, there are student's listening scores of TEM in nine years (2010, 2012-2019) from three groups, shown in Table 2. The form of TEM, in the second stage (2010-2015) is different from that in the third stage (2016-2019), as well as points, percentages and total points, which are mentioned in 2.2.2. Therefore, year's scores should be adjusted according to their total points and percentages.

2.3.2. Data Adjustment

In the study, every year listening points, dictation points are named LP2 and DP2 in the second stage, and LP3, and DP3, in the third stage, respectively, according to those shown in Table 1. Two parts percentages are 15% and 15% out of 100% in the second stage, while they are 20% and 10% out of 100% in the third stage. The total points and percentages of two parts are the same before and after 2016. Because the weigh and percentage of listening parts are different in two stages, the primitive data need to be adjusted and the study makes following steps to adjust listening score in order to find the trend of listening score and compare them in three groups from 2010 to 2019.

Firstly, counting per year weighed score of students' listening scores

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202205_5(5).0010

Every year weighted score (WS) is gotten by following formula:

$WS = \frac{LP2 \times 15\% + DP2 * 15\%}{15\% + 15\%}$ (in the second stage)

$WS = \frac{LP3 \times 20\% + DP3 * 10\%}{20\% + 10\%}$ (in the third stage)

Secondly, calculating the weigh mean (WM) of total scores of two parts in the two stages

So, the weighed mean of two parts is WM2 and WM3, respectively in two stages.

$WM2 = \frac{15 \times 0.15 + 15 \times 0.15}{0.15 + 0.15} = 15$

$WM3 = \frac{20 \times 0.2 + 10 \times 0.1}{0.2 + 0.1} = 16.67$

Thirdly, Calculating the percentage of score per year in the students' listening section

Adjusted score (AS) of every year is gotten by Formula I in the second stage and by Formula II in the third stage.

Formula I: $AS = \frac{WS}{WM2}$

Formula II: $AS = \frac{WS}{WM3}$

2.3.3. Data Analysis

The primitive scores of listening part and dictation part and their WS and AS in every year are shown in Table 2. And the figure 1 describes the trend of AS from 2010 to 2019.

It is seen from Table 2 that scores of listening in the third stage are higher than those in the second stage, except that in 2015. Although it is said that listening and dictation parts are harder in the third stage than those in the second stage, student's scores are not lower than before and the weighted score (WS) of the third stage is also higher than that of second stage. Table 2 also shows that the scores of NWAFU are higher than those of STCs and NCs in any one of nine years. Figure 1 shows that the trend of AS rises up from 2010 to 2019. However, the AS is below 0.6 in the second stage, except 2015, while that is higher than 0.6 in the third stage for NWAFU. For STCs and NCs, that is up several percent points.

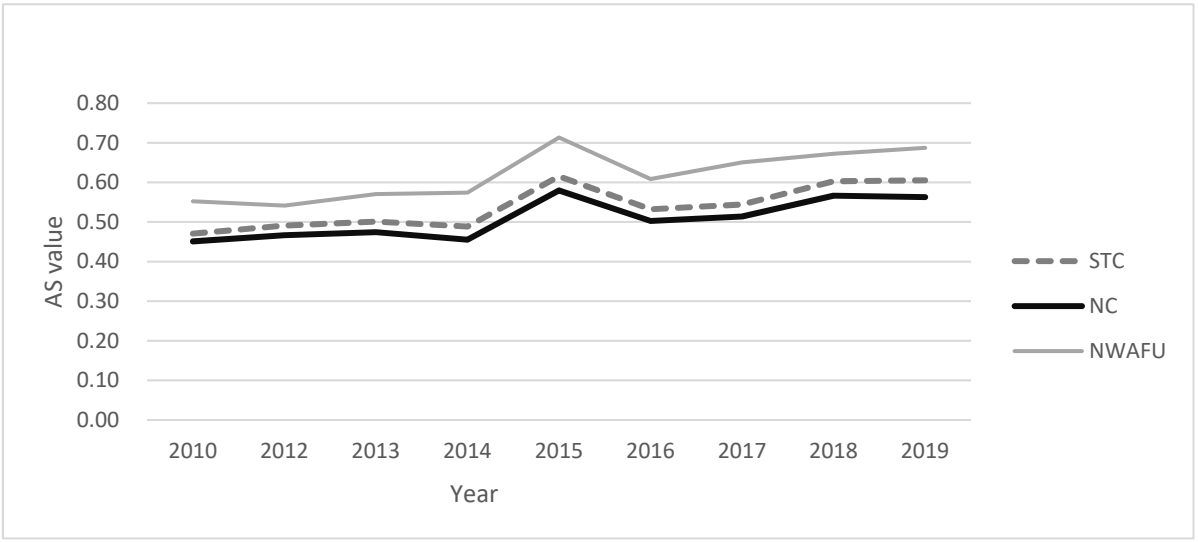

Figure 1. Nine year AS about student listening competence

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202205_5(5).0010

Table 2. Year Scores of Listening & Dictation Parts of TEM in Three Groups

	listening			dictation			WS			AS (WS/WM)		
Year	STC	NC	NWAFU	STC	NC	NWAFU	STC	NC	NWAFU	STC	NC	NWAFU
2010	9.79	9.53	10.51	4.33	4	6.06	7.06	6.77	8.29	0.47	0.45	0.55
2012	8.87	8.67	9.13	5.86	5.33	7.11	7.37	7.00	8.12	0.49	0.47	0.54
2013	9.2	8.98	9.92	5.83	5.25	7.2	7.52	7.12	8.56	0.50	0.47	0.57
2014	9.14	8.81	9.41	5.51	4.85	7.81	7.33	6.83	8.61	0.49	0.46	0.57
2015	9.64	9.32	10.51	8.82	8.07	10.89	9.23	8.70	10.70	0.62	0.58	0.71
2016	11.2	10.66	12.46	4.21	3.82	5.51	8.87	8.38	10.14	0.53	0.50	0.61
2017	11.08	10.58	12.91	5.06	4.54	6.71	9.07	8.57	10.84	0.54	0.51	0.65
2018	12.17	11.58	13.03	5.8	5.16	7.57	10.05	9.44	11.21	0.60	0.57	0.67
2019	11.68	11.01	13.21	6.9	6.13	7.95	10.09	9.38	11.46	0.61	0.56	0.69

3. Result

3.1. The Trend of Listening Score During 10 Years

As is mentioned above, before 2015 or after 2015, the AS is growing steadily. However, for most colleges, the AS cannot get the pass mark (mark=0.6, 60% of 100%). It shows that student listening ability cannot meet the requirement of TEM syllabus, especially before the third stage according to the student average score of TEM. It is an exception in 2015 that the listening scores suddenly rises up a lot, which the highest AS for three groups in their TEM history.

The trend of AS and an exceptional AS illustrate that the washback of TEM on student listening learning is strong from test factor perspective.

3.2. The Washback of TEM on Teaching Affairs in NWAFU

English major training program in NWAFU has gone through three changes after 2010, the first change is from 2009 to 2013 (called TP1), the second change is from 2014 to 2018 (called TP II), and the third change starts from 2019.

Before adjusting the program, a lot of work had been done, like surveying other colleges, conducting questionnaire several times, asking advice from the superior department, revising the scheme again and again, and last making a final decision. However, among those work, referring student the TEM score is more important one.

The study explores listening course change in TP1 and TP2 and some information was gotten from Mr. Yang, the director of teaching and research office of English major, and teachers. They said students made little progress on listening during a couple years. And their lower TEM listening scores urged the training program to be adjusted. Some curricular may be omitted and more helpful courses would open.

The listening courses for freshmen and sophomores are shown the following Table 3 in first and second periods. Because students have not taken the TEM test, the third one is left out in the study.

Table 3 shows that there are a big change for listening courses in the TP2. Listening I-IV are replaced by comprehensive I & II. For sophomores, they have news and science listening and the total listening class hours were also added to 192 hours, compared with 128 hours in the TP1. Moreover, the test training hour is also added 4 hours, from 8 to 12 hours.

What does Training Program change influence on learning outcome? Firstly, the study turns back to TEM listening score of NWAFU. The weighed score (WS) of listening is 8.56 in 2013, but it is 8.61 in 2014. The WS added up 0.05, but for STCs and NCs, both of the WS decrease from 7.52 to 7.12, and from 7.33 to 6.83. The drop points are 0.4 and 0.5, respectively. However, the trend of WS of NWAFU is the same with that of STCs and NCs from 2010 to 2019, except from 2013 to 2014. And in 2014, the training program was revised and listening courses for sophomores were added including class hours and test training hours.

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202205_5(5).0010

Table 3. Listening Courses for Freshmen and Sophomores in NWAFU

Training program	Period	Students	Course Name	Class Hours
TP1	2008-2013	freshmen	listening I	32
			listening II	32
		sophomores	listening III	32
			listening IV	32
total			128	
testing training hour for listening				8
TP2	2014-2018	freshmen	comprehensive I	32
			English phonetics	32
		sophomores	comprehensive II	32
			news listening	32
			science listening	64
total			192	
test training hour for listening				12

From the case of NWAFU, some results are gotten. Washback of TEM facilitates training program changes. Moreover, new change of listening course promotes students got better score in TEM listening part. It is interesting that the biggest gap of WS between NWAFU and STCs or NCs is also in 2014 in the second stage.

So, the answer for the second research question may be gotten. Washback of TEM strongly influences on NWAFU teaching affairs. It facilitates courses are adjusted to improve students' English level, especially on listening course arrangement. And it also promotes the change of training program. On the contrary, the change also helps students from NWAFU get better scores in the TEM listening part.

3.3. The Washback Effect of TEM on Test Making

Abovementioned TEM history in China, it has three changes from 1990. In the first stage, before 2004, the first part of TEM was writing. However, such part order was proven not to be suite for major test regular order and students could not exert their abilities very well during the test. After all, writing part is the difficult and main part for students and its weigh is bigger than any other parts of TEM. Most students could not keep calm when they face the difficulty question at the first sight during taking the test. Although there is no evidence to illustrate that the order change made students TEM score change a lot, several interviewers who teach English major students more than 20 years said that the order change really facilitate students to get better scores. They said more and more English major students were willing to take the test even if it was not a requirement for them before 2004.

In fact, TEM commitment and test makers often summarize TEM result and students score every time. The test content and form change in the third stage is a better example to illustrate that.

As mentioned above, TEM was reformed in 2016 and it is a big change, compared to two pervious changes, especially in listening part.

In the study, the listening score includes those of listening and dictation part of TEM. From Table 1, in the second stage, listening and dictation have the same points and weigh. They are 15 points and 15%, respectively. However, students score of dictation is never over 9 points (60%, 9 out of 15 points, a pass mark) from 2010-2015. Even if it is the highest in 2015 (it is 8.82 for STCs and 8.07 for NCs) in the second stage, they are not over 9 points, and only NWAFU

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202205_5(5).0010

is over 9 and it is 10.89. Whereas, for listening part of TEM, students score is over or almost passes 9 for three groups in the second stage. However, the AS (adjusted score, it is mentioned above in 2.3.1) is never over 0.6 in the second stage, except in 2015, in which it is widely believed that it is the simplest TEM in the second stage.

In the third stage (from 2016-2019), the weighed score (WS) is bigger than that in the second stage for all three groups, except in 2015. Moreover, the AS is also higher than that in the second stage, except in 2015. For NWAFU, the AS is over 0.6 after 2016, and it is higher and higher, due to the reformed TEM. In fact, the same changes also happen in other two groups.

In a word, the TEM has a big change on the content and form in 2016, called reformed TEM, especially listening and dictation part reform. The AS is hardly over 0.6 before reformed TEM. However, the AS of NWAFU is over 0.6 after 2016 and the AS of STCs is equal or over 0.6 in 2018 and 2019. The change of AS during 9 years illustrates the washback of listening part also strongly influences on test content and form. Test designers often think about the order of every question, how to arrange the difficulty of questions, what kind of form and content can really examine student's abilities through the test results. Furthermore, this washback is positive for test designers. They often receive the feedbacks from the examiners, teachers and students, and decide to how to design next test paper.

4. Discussion and Conclusion

The study explores listening score of TEM in three groups from 2010 to 2019. During this period, TEM was reformed and changed a lot on listening content and form in 2016. When talking about the listening parts of TEM in the study, it refers listening part and dictation part of TEM. Students' listening score is the sum of the score of listening part and the score of dictation part in the study. Because points and weigh of two parts are different before and after reformed TEM, the study adjusts students' listening score according to the weigh and point percentage. And next, the study calculates the WS, WM, and AS. These steps are done in order to describe how the trend of listening score goes on. From the Figure 1, it is seen that the trend of listening score in three groups rises up. And the score is higher after than before reformed TEM.

The listening score fluctuates little before or after 2015 in three groups but it abruptly rises up a lot in 2015, and reaches the top point during 10 years. It is a reason that, in 2015, the test questions of listening parts of TEM are not as difficult as that in other years for students. So, the study omits the score in 2015 when talking about the trend of listening parts of TEM in from 2010-2019.

Obviously, from figure 1, the fluctuation trend of listening score in NWAFU is the similar with that in STCs and NCs in most years. However, in 2014, when the listening score decreased in other two groups, the score in NWAFU increased a little. The study found one of reasons. It is due to the training program changed in NWAFU and listening curricular were adjusted. The listening class hours are added 68 hours and the test training hours added 4 hours. And the director of NWAFU said that it is a direct reason that students' listening score of TEM is so low that they had to consider to change listening courses and add more class hours. Moreover, the listening course content is mostly similar to the news from BBC or VOA. On one side, students are interested in current news, and on the other side, they can train their listening test skills in class. Many a teacher said the new training program on courses is better than the old one, especially on listening course adjustment. The washback effect of TEM on course arrangement is strong. It prompts NWAFU training program of English major to become more suitable for students learning. Teachers said they were proud of students' test results on TEM in 2014.

Some students who were also interviewed said that they learn about TEM listening parts are difficult for a tester, especially dictation part and they spent more time on practicing dictation.

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202205_5(5).0010

Some students said they did much more previous test papers about listening part and thought that it was hard for them to get good scores on this part and the washback strongly push them to practice more.

From mentioned above, the finding is the washback strongly affects the arrangement of listening course of English major in NWAFU and student's practice on the part. Moreover, the effect is positive. It is in accord with the finding of He (1998). He said that the washback of test on grammar teaching is positive and it is also in line with Qi 's finding that the washback can prompt student's comprehensive ability. Here, the washback improves students' English listening ability.

Before reformed TEM, student listening score from three groups hardly reached the pass mark and could not got 60% of total listening scores, and the AS of NC had never been passed 0.5 from 2010-2014. However, when 2015 omitted, the AS of every group is higher than 0.5 after reformed TEM. And the AS raised a lot for every group, compared with before reform. The main reason is test form and content are changed, especially in listening parts (listening part and dictation part). Before reform, the weigh score of dictation part is 15% of total scores of TEM and that of listening part is the same. But after reform, the dictation weigh is 10% and listening weigh is 20%. Moreover, the contents of listening parts are also different. The detail is shown in Table 1. It is clear that test form and contest changes affect the student test results. And lower scores of students in listening parts for a long term makes test makers do some changes. It is the wash back effect of TEM on these changes. From the feedbacks of teachers, these changes are more suitable for evaluating student listening abilities, and the students said they had more confidence to take the test when they comparted after reformed TEM test papers with before ones. The experts and society also give positive response for the TEM reform.

From above mentioned, the finding is that the washback of TEM positively influences on test content and form changes. It is in consistent with the findings of Shih (2007) that the washback effect of test should also be considered from test factors. And it is also in line with what Cao Qin (2009) found that the washback of CET listening part on student's learning is positive.

From all above, the washback of TEM listening parts on student learning and test making is strong and positive. The washback prompts students to spend more time on listening practice and listening class hours are also increased for students. The washback gives a strongly effect on listening course arrangement. Meanwhile, the washback also impacts test form and content changes positively.

Although the study is simple and narrow, it explains that the washback of TEM on school teaching affairs and test making which is lack in previous studies. The study also suggests that teaching affair office should consider which courses are suitable for improving student listening ability and how to arrange those, and the Test Committee should consider how to make the washback have a positive effect on teaching and learning. The further study will explore more details about washback of other TEM parts on student learning and an empirical study will be conducted.

Acknowledgments

Firstly, the study has been approved by the project "The Study of the Construction of Oral CAF Evaluation System Based on English Corpus" (No. XGH21053) from Shaanxi Higher Education Society and by the teaching project "The Study of the Optimization and Evaluation of Oral CAF Indexes Based on English Corpus" (No. JY2103201) from Northwest A&F University. The authors thank the departmental Office for the support. Secondly, the authors also thank co-workers for their helps in the study. At last but not least, the authors appreciate what assistants and other persons have done for the study.

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202205_5(5).0010

References

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

[2] Andrews. S., J. Fullilove, Y Wong. (2002). Targeting Washback—a Case Study. System, 30, 207-223.

[3] Bailey, K. M. (1996). Working for Washback: A Review of the Washback Concept in Language Testing. Language Testing, 13(3), 257 –279.

[4] Buck, G. (1988). Testing Listening Comprehension in Japanese University Entrance Examinations. JALT Journal, 10, 17-42.

[5] Cao, Q. (2009). The Washback Effect of the Reform of College English Test Band 4 Listening Test on College English Listening Teaching (Unpublished master thesis). Southwest University, China.

[6] Dong, J.H. (2009). The washback effect of new CET band-4 reading upon college English teaching and learning] (Unpublished master thesis). Shangdong Normal University, China.

[7] Enfari, 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.

[8] Gu, X.D. (2004). Positive or negative? —An empirical study on the backwash effect of CET-4 and CET-6 on college English teaching in China (Unpublished Doctoral Dissertation). Chongqing University

[9] He, C.Z. (1998, July), The Backward Effect of Professional Grammar Teaching and TEM-4 and TEM-8 Tests. In Zu & Shen(Eds), the Theory and Practice of English Language Testing. the Proceedings of Foreign Language Education (pp. 230-238), Shanghai, China: Shanghai Press.

[10] Hong, Y., & Yi, l. (2016). Positive and Negative Backward Effects of TEM4 and TEM8 Listening Tests on Listening Teaching of English Majors. Foreign Language Education and Rranslation Development Innovation Research, 5:125-127.

[11] Huang, D.Y., & Yang, B.. (2002). Overview of research on the backwash effect of language testing. Foreign Language Teaching and Research, 34(4), 288-293.

[12] Huang, H.L. (2004), The Backward Effect of CET-4 Oral Test on Oral English Teaching and Its Implications. Journal of Shaoguan University, 2 (11), 133-136.

[13] Hughes, A. (1989). Testing for Language Teachers. Cambridge: Cambridge University Press, 1989.

[14] Hughes, A. (1993). Washback and TOFEL 2000 (Unpublished Manuscript). University of Reading.

[15] Kong, L. (2014). Testing Validity and Washback Effect of "Grammar and Vocabulary" Questions in TEM4 (Unpublished master thesis). Nanjing Normal University, Nanjing, China.

[16] Li, X.J. (1997, September 30). The Science and Art of Language Testing. Hunan Education Press, pp. A2.

[17] Liu, Y. (2005). A study of TEM4 Washback Effect (Unpublished master thesis). Shangdong Normal University, Jinan, China.

[18] Liu, B., & Hu, F. (2018). A Study on Validity of CET-4 Reading Test ---Taking the Reading Task of TEM4 (2016). New Outline Sample Questions as an Example. Technology Enhanced Foreign Language Education, 179: 70-75.

[19] Meng, Q., & Li, J. (2019). A Case Study of Grammar and Vocabulary in TEM-4 ---New Thoughts on English Grammar Classroom Teaching. Foreign Language Testing and Teaching, 3, 34-43.

[20] Prodromou, L. (1995). The Washback Effect: From Testing to Teaching. ELT, 4(1), 13 -25 .

[21] Qi, L.X, (2011). The Washback Effect Theory and Empirical Research of Language Testing. Theory of Foreign Language Teaching, 10(4), 23-28.

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202205_5(5).0010

[22] Shih. (2007). A New Washback Model of Students' Learning. The Canadian Modern Language Reviews, 64(1), 135-162.

[23] Shohamy, E. (1996). Test impact revisited: Washback effect over time. Language Testing, 13 (3), 298-317.

[24] Wang, J., & Xia, L (2018). A Study on the Correlation Between Students' English Vocabulary, Autonomous Learning Ability and TEM-4 Scores. Journal of Tianjin Radio & TV University, 22(3), 44-49.

[25] Watanabe, Y. (1992). Washback Effects of College Entrance Examination on Language Learning Strategies. Japan Association of College English Teachers Bulletin, 23: 175-193.

[26] Wu, J. K. (2019). A Study on the Washback Effect of reformed CET-4 on Students' English Learning. Overseas English, 4, 111-112.

[27] Xu, Q. (2014). A study washback effect of TEM (Unpublished master thesis). Shanghai International Studies University, Shanghai, China.

[28] Yan, J. L., & Yang, H. (2011). The Interaction and Washback Effect of TEM-4 and TEM-8 Tests on English Teaching -- A Case Study of English Teaching Reform in Polytechnic Colleges, Foreign Language Testing and Teaching, 2, 55-79.

[29] Yu, H. (2013). An Empirical Study on Washback of CET4 Writing to First and Second-Year College English Teaching and Learning (Unpublished master thesis). Yangtze University, Jingzhou, China.

[30] Zhang, S. Y., Huang, L., Jiang, Y., Luo, W. (2019). The Washback Effect of TEM-4 Reform on Students' Learning. Oversea English, 11, 66-69.

[31] Zhang, Y. (2017). The Reform of Listening Questions in 2016 TEM-4 and Its Enlightenment to Teaching. Education Development, 2(6), 259-206.