



Washback of High School Examinations and its effects on Iranian EFL Learners' Language Learning Beliefs

Dr. Peyman Rajabi

Assistant Professor, Department of ELT, Malayer Branch, Islamic Azad University, Malayer
Paymanrajabi2002@yahoo.com

Maryam Nazari

PhD Candidate in TEFL, Malayer Branch, Islamic Azad University, Malayer
maryamnazari17@yahoo.com

Abstract

The phenomenon of how tests influence teaching and learning is commonly described as “washback” in language instruction. Literature indicates that testing washback is a complex concept that becomes even more complex under different interpretations of the washback phenomenon on teaching and learning. Based on recent studies, washback effect can be positive or negative. In the present study some definitions of washback and its two major types were introduced, and also the effect of test on teachers and learners was indicated. Something that should be considered is the validity of the test, which influences the effect that the test has on teaching or learning. However sometimes there are other factors which do not let the test have its right effect on teaching and learning processes. Now it is important to know how to promote washback. Recently the field of ESL/EFL has begun to recognize the importance of students' role in the nature of language instruction. For teachers to be able to provide effective instruction, it is necessary to be aware of the abilities of their students, make correct decisions about what they are going to do during the course of instruction, the way of presenting the materials and also to evaluate themselves and their students correctly in order to succeed in teaching. This study focused on the introduction of learners' beliefs and their influence on teaching and learning process, and also the relationship that exists between students' language learning beliefs and the washback effect of high school English examinations.

In conclusion, a framework for measuring washback effect requires a more comprehensive model. Language testing is not limited to classrooms, and its consequences are not only educational, but also social and political. It is clear that language testers must accept responsibility for all those consequences which they are aware of.

Keywords: washback, positive washback, negative washback, belief, washback effect.



Introduction

The importance of testing and its effects on teaching and learning processes is a fairly recent phenomenon. This is especially true for teaching and learning English as a foreign language in our country.

According to Cheng (2005), language tests are used for a variety of purposes, including:

- making inferences or predictions about test takers' language abilities or to make predictions about their capacity for using language to perform further tasks in contexts outside the test itself;
- making decisions (e.g., selection, diagnosis, placement, progress, grading certification, employment) about test takers on the basis of their test scores, which can inform the decision-makers about their ability or their capacity for language use in non-test situations. (p. 25)

Testing is important because it can provide information, motivation, and reassurance for teachers, learners, and administrators. If testing is done for these purposes, it may often be harmful in different ways. The effects of testing are the result of considering the test procedures as ends in themselves, rather than instruments to the achievement of some goals.

Statement of the Problem

In recent years many language researchers and educators have considered various levels of washback effect on foreign language instruction. There is evidence to suggest that examinations may have washback effect on teaching and learning (Alderson & Wall, 1993). According to Bachman (1990), positive washback occurs when the testing procedure reflects the skills and abilities that are taught in the course. However, when "there is little or no apparent relationship between the types of tests that are used and instructional practice", negative washback occurs (Bachman, 1990, p. 283). Despite numerous studies regarding the testing effect on EFL teaching and learning in different contexts, research is still lacking on the washback effect of high school examinations and learners' beliefs, particularly in Iran. Thus, this study was designed to investigate the scope of washback effect of examinations in English education in Iran high schools and to find out if it has any relationship with learners' beliefs.

Review of Literature

It has been for about four decades that researchers have come to this conclusion that the tests which are administered during the course of instruction have an undeniable effect on the process of teaching and learning. Some studies conclude that no simple washback effect occurs (Alderson & Hamp-Lyons, 1996), whereas others find powerful determiners of language testing toward classroom teaching (Herman & Golan, 1993; Hughes, 1989).

Washback and the impact of tests have recently become a major area of study within educational research, and language testing in particular. For both teachers and students there is a natural tendency to make an appropriate relationship between their classroom activities and the test. Sometimes, the test is very important to students and teachers. Therefore, the importance of washback or the effect of test on teaching and learning process increases. The validity of the test influences the quality of the effect that a test has on teaching and learning, which means a good test will have good effects and a bad test will have bad effects. So, there must be a close relationship between the test content and the content of teaching materials.



Considering this relationship, there may be positive or negative washback, although in producing negative washback, sometimes the factors other than the test itself may be influential. Teachers should consider some points in order to promote beneficial washback.

Bachman & Palmer (1996, pp. 30-31) considered micro and macro levels for washback, although they included the influences on individual teachers under the micro category. The language learners as well as the other participants affected by washback may be influenced by official information about a test prior to its administration (including advertising materials from the test publisher, existing test preparation booklets, etc.), or by folk-knowledge (such as reports from students who have taken earlier versions of the test). They may also be influenced by several sources of feedback following the administration of the test. These would include the actual test scores provided by the exam scoring service, feedback from the test takers (what was easy or difficult, what seemed fair or unfair, unexpected item types, unfamiliar instructions, etc.), feedback from the proctors if the test was administered locally, and feedback from the teachers in reaction to the students' scores. The information might be officially supplied (via score reports and information bulletins), inferred, or even imagined.

Herman & Golan (1993) conducted a survey among two groups of teachers from two different kinds of schools where test scores had increased or test scores had decreased or remained the same. Finally, they came to these important conclusions that without observational data nobody knows how tests influence teaching process, how the tests influence planning, how much time is required to prepare the students for test taking, and what kind of attention is given to those subject areas that are not covered in tests. Therefore, survey data are useful, but insufficient, for understanding language teaching washback.

As stated, washback is known as the effect of testing on teaching and learning. It can be harmful or beneficial. According to Hughes (1989), "if a test is important, then preparation for it can come to dominate all teaching and testing activities. And if the test content and testing techniques are at variance with the objectives of the course, then there is likely to be harmful washback" (p. 1). For example, if teacher tests the skill of writing only by multiple-choice items, then there is great pressure to practice these items rather than practice the skills of writing itself, which is clearly undesirable.

The term washback is neutral because the influence of a test may be either positive or negative in nature. That is, a poor test leads to negative washback, while a good test will have effects understood as positive.

Alderson & Wall (1993) mentioned some of the negative effects of tests as follow:

Narrowing or distortion of the curriculum, loss of instructional time, reduced emphasis on skills that require complex thinking or problem solving, test score pollution, and meaning gains in test scores without a paralleled improvement in actual ability in the construct under examination. (p. 115)

However, Swain (1985) and Alderson (1986), like some other researchers emphasized the potential positive aspects of test influence. They focused on the development of tests, which through constructive washback will have informative effects on language curriculum.

Hughes (1993) suggested a framework: "In order to clarify our thinking about backwash, it is helpful, I believe, to distinguish between participants, process, and product in teaching and learning, recognizing that all these may be affected by the nature of a test" (p. 2). In Hughes' framework, participants include language learners and teachers, administrators, materials developers, and publishers, "all of whose perceptions and attitudes toward their work may be affected by a test" (ibid). The term process refers to "any actions taken by the participants



who may contribute to the process of learning” (ibid). According to Hughes, such processes include materials development, syllabus design, changes in teaching methods or content, learning and/or test-taking strategies, etc. Finally, in this framework, product refers to “what is learned (facts, skills, etc.) and the quality of learning (fluency, etc.)”. (ibid)

The question of why some EFL learners rate higher in English language achievements than others has been at the center of much research and controversy for many years. Recently the field of ESL/EFL has begun to recognize the importance of learners’ role in the nature of language instruction. For the teachers to be able to provide effective instruction, it is necessary to be aware of the abilities of their students, make correct decision about what they are going to do during the course of instruction, the way of presenting the materials and also to evaluate themselves and their student correctly in order to make a successful teaching.

Numerous scholars have written on the topic of learner, learner characteristics, behaviors, beliefs and different factors which may influence a good learning process. Recent researches have shown that if there is a kind of relationship between teachers and students, learning will be facilitated (Darby, 2005).

Emphasis must be placed upon learners’ beliefs as what learners think about the world, their place within it and also their values. There seems to be kind of relationship between the belief and behavior of learners, i.e. difference in beliefs causes changes in behavior in the classroom.

Usually in school systems there are mid-term and final examinations. As educators believe they have some effect on the education process, and learners’ beliefs are not exception in this case, because learners are highly influenced by their beliefs, which are their values, their views of the world, and their conception of their place within it. According to Williams and Burden (1997), beliefs are culturally bound, are formed early in life, and are resistant to change.

Nespor (1987) claimed:

Beliefs are closely related to what we think we know but provide an affective filter which screens, redefines, distorts, or reshapes subsequent thinking and information processing. (p.19) According to Hassaskhah (2006), virtually all learners, especially older ones, have strong beliefs and opinions about how their instruction should be delivered. Several studies have been done on this concept, and they have revealed that learners’ progress was negatively affected by an instructional approach which was not consistent with their beliefs about the best way for them to learn.

Subjects

A total number of 120 students participated in this study. They were selected randomly from some high schools in different parts of Kermanshah. All the students had studied English for at least three years. They were all high school students and ranged between 13-18 years of age. Their final English score ranged between 13-20 and the mean was 18.

Regarding the gender of the participants of high school students, it should be mentioned that all of them were female. And regarding their grade, 30% were studying in the first grade, and the number of those studying in the second and third grade were the same (Both of them 35%). The subjects who participated in this study were mainly from first grade and natural sciences (Uloom-e-Tajrobi). Nearly 28% were from mathematics (Riaziat) field of study, and about 12% of them were studying Uloom-e-Ensani.

Instrumentation

The instrument used in this study was the BALLI (Horwitz, 1987). It is a 34-item questionnaire, containing statements related to the following five areas: i) foreign language aptitude, ii) the difficulty of language learning, iii) the nature of language learning, iv) learning and communication strategies; and, v) motivation and expectations. Respondents are required to rate their agreement to each statement on a Likert-type scale from 1 (strongly agree) to 5 (strongly disagree). The statements on a Likert scale should be either expressing a positive/favorable or a negative/unfavorable attitude towards the object of interest. Although the Likert-scale was originally developed to measure attitudes, its scope has been extended to wider cognitive and affective variables, including beliefs.

While the BALLI is a widely used and recognized questionnaire in research on learner beliefs, it includes a prescriptive set of statements in which respondents mark their degree of agreement. Furthermore, as with any survey instrument, there is a chance that respondents may misunderstand certain items due to either their own limited language proficiency or the subjective nature of the item (e.g., Item 15).

A questionnaire was used in order to estimate the degree and scope of learner beliefs on high school examinations in Iran. In addition, the washback questionnaire was used to obtain data about washback effect of tests.

Procedure

The necessary data were collected through using two questionnaires. One of them was about washback effect of examination and the other questionnaire was about learners' language learning beliefs. The source of the last questionnaire was an article written by Huang (2006). Two questionnaires were administered to 120 learners in some high schools in Kermanshah to elicit their views on the effect of washback of high school examinations and LLLB. The participants were provided with the questionnaires either in their classes or by the researcher who contacted them through some language institutes. All participants completed the questionnaires outside of class and returned them either to their teachers, who forwarded the questionnaires on to the researcher, or directly to the researcher herself.

Having collected the data through the questionnaires, the researcher analyzed the data to find out whether the washback of high school examinations has any statistically significant effect on Iranian EFL learners' language learning beliefs. This analysis was done through the SPSS program.

Results

Design

This study had a questionnaire design. The subjects were randomly chosen. The study was conducted on a group of 120 female students. The Horwitz' BALLI questionnaire for measuring learners language learning beliefs and a washback questionnaire were used.

Data Analysis

To achieve a reasonable answer to the research questions, descriptive statistics on the variables were tabulated. As mentioned earlier in chapter one, the research hypothesis is as follows:

The washback of high school examinations does not have any statistically significant effect on Iranian EFL learners' language learning beliefs.

Table 4.1- Descriptive statistics of washback effect questionnaire

variable	Descriptive statistics							
	Valid N	Mean	Median	Minimum	Maximum	Variance	Std. Dev	Standard Error
Wash back	120	3.41	3.46	2.38	4.54	0.20	0.45	0.04
General	120	3.17	3.17	2.17	4.00	0.13	0.36	0.03
Test content	120	2.91	2.94	1.88	4.00	0.19	0.44	0.04

Table 4.1 presents the data on the mean and standard deviation of the students (120 on the whole) regarding their views about different factors of washback effect. These factors are categorized into three subscales that are washback, general, and test content.

According to Table 4.1, the subjects showed high mean (higher than 3), in washback and general. However, in test content they showed low mean (lower than 3). Therefore, it shows that they believe test content does not have so much influence on washback effect. They showed low standard deviation (lower than 1) regarding their ideas on the washback effect of high school English examinations. It seems that high school students as a whole agreed on the washback effect of English high school examinations.

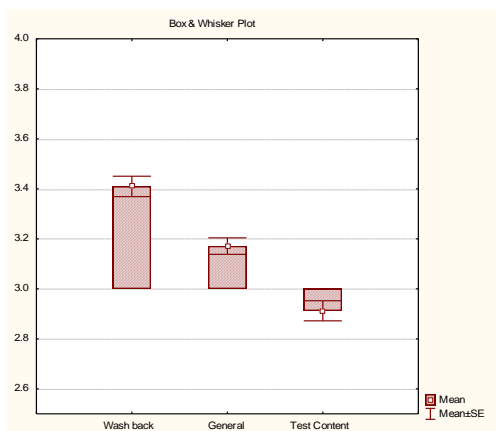
Because the number of items for each factor is not the same, each person's mark is divided based on the number of items in each factor in order for the factors to be comparable. And the result is a number between 1 to 5.

Table 4.2- Descriptive statistics of learners' belief questionnaire

variable	Descriptive statistics							
	Valid N	Mean	Median	Min	Max	Var	Std. Dev	Standard Error
Self Efficacy	120	3.31	3.33	2.17	4.33	0.20	0.45	0.04
Perceived Value	120	3.56	3.67	1.78	4.89	0.40	0.63	0.06
Language Aptitude	120	3.06	3.14	1.57	4.43	0.42	0.65	0.06
Formal Structural Studies	120	3.45	3.43	2.14	4.71	0.36	0.60	0.05
Others	120	3.35	3.33	1.83	4.50	0.32	0.57	0.05
Learner's Belief	120	3.35	3.47	2.57	4.24	0.17	0.41	0.04

Table 4.2 illustrates students' views on the correspondence between different factors of learner beliefs and foreign language learning. It seems that students agreed on such a correspondence. They showed a high mean and low variation in their ideas about it. And if we want to compare the mean of different variables, we conclude that perceived value factor has the highest mean among others and language aptitude factor has the lowest mean.

Figure 4.1- Box & Whisker Plot



As it is presented graphically in Figure 4.1, washback factor has the highest influence and test content has the lowest influence. And washback, general and test content are sequenced respectively according to their influence on washback effect.

Figure 4.2- Box & Whisker Plot

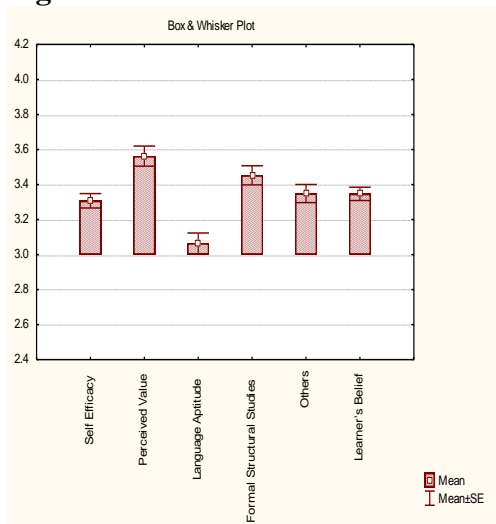


Figure 4.2 indicates that the learners' beliefs components can be sequenced based on their influence on foreign language learning. Subjects showed high agreement on all these components because the mean values for all are higher than 3.

Conclusion

Summary of findings

In the present study, the results obtained from the questionnaires that were administered to 120 high school students indicated that in general, students agreed on the washback effect of English high school examinations. And also they agreed with the correspondence between different factors of LLLB and foreign language learning.

According to the learners' answers to the items of the questionnaires, among the different subscales of both questionnaires, for the belief questionnaire the perceived value has been highly accepted by the students as being influential and for the washback effect one the washback subscale was more influential than the others.

Now the important point is to distinguish the relationship between the two variables, washback effect and learner language learning beliefs. In order to show the degree of relationship, the correlation between both variables is measured. Table 4.3 in Appendix A indicates the correlation between learner beliefs in relation with each of the components of washback effect.

As it is shown, the first two components, washback and general are in relation with learners' beliefs, but test content lacks such relationship. The correlation square shows that 0.08 of the changes in the dependent variable that is learners' beliefs is due to washback and 0.15 of that is because of general component, and 0.01 of the changes is due to test content. In order to indicate that whether these results are the same if we consider the whole statistical society, the test for meaningfulness of correlation was applied. Because the amount of P is less than 0.05, the Ho is rejected and we conclude that there is a kind of relationship between learners' belief and washback effect.

Measuring the correlation between both variables showed that learners' beliefs has a kind of relationship with washback and general subscales. Of course, after considering all the aspects, we came to the conclusion that learners beliefs and washback effect are in relationship with each other.

Figure 4.3 in Appendix A shows the correlation between washback and learner beliefs. As the washback line goes higher and higher, the learners' beliefs line goes higher too. Therefore, there is a relationship between washback and learners' beliefs.

Table 4.3 also indicates that there is a relationship between learner beliefs and general component. Figure 4.4 shows this correlation. Again the higher the line of general, the higher the line of learners' beliefs. So, there is a positive relationship between these two.

However, because P is less than 0.05, test content does not have such a relationship with learners' language learning beliefs, so the higher the line of test content, the lower that for LLLB.

Table 4.4 in Appendix A shows the degree of relationship between different components of washback effect and different components of learners' beliefs. As it is indicated washback has meaningful and positive correlation with all the components of learners' beliefs but not with self-efficacy. And the correlation between general and different components of learners' beliefs is the same too. Again it does not have relationship with self-efficacy. But test content has negative correlation with those components because P here is higher than 0.05.

Table 4.5. - Multiple Regressions

MULTIPLE REGRESSION RESULTS:

Variables were entered in one block

Dependent Variable: Learner's Belief

Multiple R: .4976

Multiple R-Square: .2476

Adjusted R-Square: .2282

Number of cases: 120

F (3, 116) = 12.729 p < .00

Standard Error of Estimate: .3641

Intercept: 2.153 Std.Error: .3378

t (116) = 6.3749 p < .00

Summary Statistics; DV: Leamer's Belief (NazariFinalData)	
Statistic	Value
Multiple R	0.50
Multiple R ²	0.2477
Adjusted R ²	0.2284
F(3,116)	12.73
p	0.00
Std.Err. of Estimate	0.36

Because each of the three components of washback effect does not have separate influence on learner beliefs, the effect of all of them together through multiple regressions was measured. Table 4.5 shows that almost 25% of the changes in LLLB are due to the three components of washback. According Table 4.5, in spite of the positive correlation between the two variables in this sample, because the amount of P is less than 0.05 we conclude that the result is the same in the whole statistical society.

Table 4.6. - Regression Summary for Dependent Variables: Learners' Belief

Regression Summary for Dependent Variable: Leamer's Belief (NazariFinalData)						
R= .49766479 R ² = .24767025 Adjusted R ² = .22821344						
F(3,116)=12.729 p<.000000 Std.Error of estimate: .36417						
	Beta	Std.Err. of Beta	B	Std.Err. of B	t(116)	p-level
N=120						
Intercept			2.15	0.34	6.37	0.00
Wash back	0.28	0.10	0.26	0.09	2.76	0.01
General	0.33	0.09	0.38	0.11	3.59	0.00
Test Content	-0.33	0.09	-0.31	0.09	-3.63	0.00

Table 4.6 illustrates that if all the three components of washback are considered together, the test content factor is not meaningless any longer. As a result, all three have positive correlation with learners' beliefs. In order to show this, the following formula can be used:
 Learners' beliefs = $b_0 + (b_1 \times \text{washback}) + (b_2 \times \text{general}) + (b_3 \times \text{test content})$

Figure 4.5. - Students' View on Washback Effect Factors Questionnaire

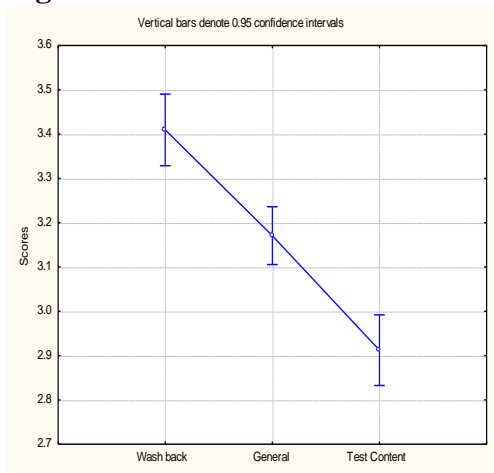


Figure 4.5 shows the subjects' view on the washback effect components. The repeated measure analysis (R.M.A.) was carried out to show that the same result exists even if despite this sample the whole statistical society is considered. Figure 4.5 shows the following results:
 1. There is a difference among the washback effect components. The score of washback is the highest and for test content is the lowest.

2. Regarding the mean of the scores in the three components, we can conclude that while students show agreement in their views on washback and general, they have different view on test content.

Table 4.7. - R.M.A. Analysis of Variance

Repeated Measures Analysis of Variance (NazariFinalData) Sigma-restricted parameterization Effective hypothesis decomposition					
Effect	SS	Degr. of Freedom	MS	F	p
Intercept	3604.46	1	3604.46	11300.8	0.00
Error	37.96	119	0.32		
R1	14.85	2	7.42	72.06	0.00
Error	24.52	238	0.10		

Table 4.7 presents data based on the R.M.A. of variance. As the F- ratio shows there is a difference between students' views on washback, general and test content. And because the amount of P was less than 0.05 the null hypothesis was rejected.

Table 4.8. - Approximate Probabilities for Post Hoc Tests

Tukey HSD test; variable DV_1 (NazariFinalData). Approximate Probabilities for Post Hoc Tests Error: Within MS = .10301, df = 238.00				
Cell No.	R1	{1}	{2}	{3}
1	Wash back	3.4096	3.1708	2.9123
2	General	0.00002	0.00002	0.00002
3	Test Content	0.00002	0.00002	0.00002

A Post Hoc analysis was applied to account for determining the differential factors. Here the factors are compared two by two. So we conclude that the score of washback is higher than general and general is higher than test content. (Table 4.8)

Figure 4.6. - Students' View on Learners' Language Learning Beliefs

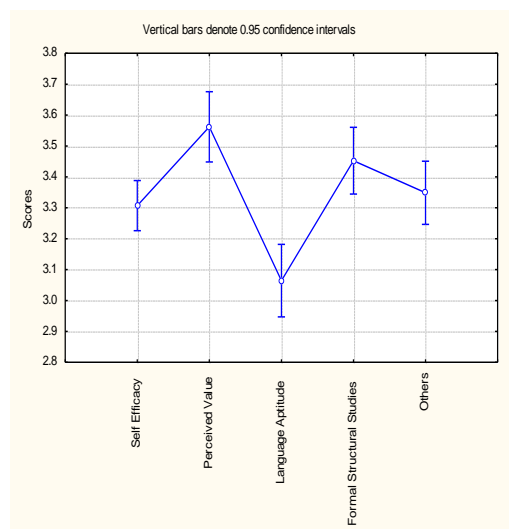


Figure 4.6 presents graphic representation of data as the subjects' view on the LLLB components. The R.M.A. analysis was carried out to indicate that if the whole of the

statistical society was considered, again the same result would be shown. This figure indicates the following results:

1. There is a difference among the components of LLLB. The mean value for perceived value was the highest and for language aptitude was the lowest amount.
2. There is a little difference between the effect of self-efficacy and other components. The subjects almost showed the same pattern regarding the effect of these two on foreign language learning.

Table 4.9. - R.M.A. Analysis of Variance

Repeated Measures Analysis of Variance (NazariFinalData) Sigma-restricted parameterization Effective hypothesis decomposition					
Effect	SS	Degr. of Freedom	MS	F	p
Intercept	6720.81	1	6720.81	7822.21	0.00
Error	102.24	118	0.86		
R1	16.67	4	4.17	19.93	0.00
Error	99.52	476	0.21		

Table 4.9 presents statistical analysis carried out through R.M.A. It indicates that there is a difference between students view on the subscales of learners' beliefs.

To find out where the difference lies, a Post Hoc analysis was applied after the ANOVA (Table 4.10). A cross-comparison of the components indicates that almost for half of the cases, the differences are insignificant. As it is shown the amount of language aptitude is less than the others.

Table 4.10. - Approximate Probabilities for Post Hoc Tests

Tukey HSD test; variable DV_1 (NazariFinalData.sta) Approximate Probabilities for Post Hoc Tests Error: Within MS = .20908, df = 476.00						
Cell No.	R1	{1}	{2}	{3}	{4}	{5}
1	Self Efficac	3.3069	3.5620	3.0643	3.4524	3.3486
2	Perceived Valu	0.00	0.00	0.00	0.34	0.00
3	Language Aptitud	0.00	0.00	0.00	0.00	0.00
4	Formal Structural Studie	0.10	0.34	0.00		0.40
5	Others	0.96	0.00	0.00	0.40	

REFERENCES

- Alderson, J. C., (1986). Introductions in Language Testing. In M. Portal, (Ed.), *Innovations in Language Testing: Proceedings of the IUS/NFER Conference* (pp. 93-105). Windsor: NFER-Nelson.
- Alderson, J. C., & Hamp-Lyons, L. (1996). TOEFL preparation courses: a study of washback. *Language Testing*, 13(3), 280-297.
- Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14, 115-129.
- Backman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Backman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Cheng, L. (2005). *Changing language teaching through testing: A washback study*. Cambridge: Cambridge University Press.
- Darby, L. (2005). Science students' perceptions of engaging pedagogy. *Research in Science Education*, 35(4), 425-445.
- Hassaskhah, J. (2006). A Window on Learners' Beliefs and Their Classroom Experience. *Learning Conference (L06)*.
- Herman, J. L., & Golan, S. (1993). The effects of standardized testing on teaching and schools. *Educational Measurement: Issues and Practices* (12), 20-25.
- Horwitz, E. K. (1987). Surveying student beliefs about language learning. In A. Wenden, & J. language Rubin (Eds.). *Learner strategies in learning* (pp. 119-129). Englewood Cliffs, NY: Prentice Hall.
- Huang, Zh. (2006). Learner beliefs of language learning revisited. *Sino-US English Teaching*, 3, 3.
- Hughes, A. (1989). *Testing for language teachers*. Cambridge: Cambridge University Press.
- Nespor, J. (1987). The role of beliefs in the practice of teaching. *Journal of Curriculum Studies*, 19, 317-328.
- 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 Directions in Language Testing*. (pp. 35-46). Oxford Pergamon.
- Williams, M., & Burden, R. L. (1997). *Psychology for language teachers: a social constructivist approach*. Cambridge: Cambridge University Press.