



MJAL 7:2 Summer 2015

ISSN 0974-8741

Effectiveness of washback in language testing: analyzing the validity and reliability by Mohammad Altowaim

**EFFECTIVENESS OF WASHBACK IN LANGUAGE TESTING:
ANALYZING THE VALIDITY AND RELIABILITY**

Bios:

Mohammad Altowaim is PhD candidate at School of Foreign Languages, Sun Yat-sen University, Guangzhou, China. He has obtained the MA in Applied Linguistics from University of Sheffield, UK. Mr. Altowaim has been teaching Arabic and English for speakers of other languages in past 10 years. He is currently interested in Language Testing, Language planning in Education. His email is: mohd.altowaim@gmail.com

ABSTRACT

Washback refers to the extent to which the test impacts language learners and teachers to do things they would not otherwise do. This study aims to analyze the validity and reliability in relation to the effectiveness of washback in language testing. The researcher collected primary data for answering the research questions and testing the hypothesis. The participants of the research are English as a Foreign Language (EFL) students. Sample consisted of 50 learners. Twenty-five students had attended test-based classes, and the other twenty-five students had attended general classes. The hypothesis compared the difference between groups using independent samples t-test. The findings of the study showed a significant difference between the mean scores of two groups. Students who had attended test-based classes had significantly lower scores than the students who had attended general classes. The study found that the test influences the classification of curriculum content into important and unimportant. It also creates a fear factor of the test in students. Teachers also emphasize on the content that is relevant to the test. They are also interested in improving the overall test score of the class. In the process, they lose sight of the total picture and the broader vision of imparting knowledge and quality education.

KEY WORDS: Washback, language testing, reliability, validity, EFL



MJAL 7:2 Summer 2015

ISSN 0974-8741

Effectiveness of washback in language testing: analyzing the validity and reliability by Mohammad Altowaim

1. Introduction

Washback is a concept used in applied linguistics that refers to the extent to which the test impacts language learners and teachers to do things they would not otherwise do (Tomlinson, 2013). The concept of systemic validity is also applied to washback. Washback validity implies that the validity of the test should be measured by the extent to which the test has a positive influence on teaching. In the fields of applied linguistics and education, there is a general conception that testing influences learning and teaching. It results in being trapped in a circle. The circle revolves as follows; ‘what is assessed becomes what is valued, which becomes what is taught’ (Safa & Goodarzi, 2014).

Testing has never been a neutral process. It always has consequences because of being a differentiating ritual for students. The principle idea in washback is that examinations or tests should drive learning and teaching. Due to this, washback is also known as measurement-driven instruction. The washback exists due to its significant impact on the lives of test takers and the high authority of external testing. Consequently, testing drives the curriculum. It also drives students’ approaches to learning and teaching methods. There are two major areas of washback. The first is related to traditional, large-scale, multiple-choice test. They may have positive or negative influences on the quality of learning and teaching. The second is related to those studies where a particular examination or test has been improved or modified to exert a positive influence on learning and teaching.

It has been argued that washback may be counterproductive for the learning of students. The negative effect may also influence the writing of the students. The argument is based on the notion that goals can impact performance when there are clear directions about what to include in an essay. The instructions may give the impression to students that an argumentative essay, for example, will be evaluated on the basis of how well the position of the writer is backed up with data. The goal instructions also make an influence on the way in which teachers teach students in argumentative writing. When the goal instruction of an exam is limited to data and claims, without any consideration of counter argumentation, there can



MJAL 7:2 Summer 2015

ISSN 0974-8741

Effectiveness of washback in language testing: analyzing the validity and reliability by Mohammad Altowaim

be a considerable impact on learning and teaching, especially in exam-oriented societies such as the Chinese society (Liu & Stapleton, 2014).

The concept of washback is important due to the impact of the test on educational systems, society, and individuals. There are two levels on which the washback is operational. The first is the micro level, in which the test has an impact on individual teachers and students. The second is the macro level, in which the test has an impact on the educational system and society. Laborda et al. (2012) analyzed the effects of washback at both macro and micro level. The findings of their study are shown in Table 1 below:

Table1 washback effects

Value/ Specificity	Factors Mediating Washback	Participants	Processes	Products
<i>General Positive Washback</i>	<i>Macrocontext:</i> General improvement of speaking skills at national level	School decisions: Bilingual, additional education in English longer classes, bilingual education (primary and secondary)	Students could perform in face-to-face situations but according to subproject 2 they could also improve the computer skills	
	<i>Microcontext:</i> Instrumental motivation: International trade; student mobility, Positive parental perception	Teacher Training and new standards in teacher education. Revision of teaching diplomas Student's Motivation intrinsecal and extrinsecal Refined goals		
<i>Specific Negative Washback</i>	Teaching in high school: Exam oriented Teacher's beliefs: Centrality of test Language learning Communicative language teaching	Syllabus Specific objectives Classroom activities and interaction In-class evaluation practices		Students' level of communicative competence clearly oriented towards the test

Tayeb et al. (2014) mentions four different notions of washback. The first is the washback effect that describes the impact tests have on learning as well as teaching. The second is the measurement-driven instruction that emphasizes that tests should drive the learning. The third is curriculum alignment that describes the relationship between the testing and the teaching



MJAL 7:2 Summer 2015

ISSN 0974-8741

Effectiveness of washback in language testing: analyzing the validity and reliability by Mohammad Altowaim

syllabus. The fourth is systemic validity that results in the integration of tests into the system of education.

The concept of washback was first introduced in 1993 by Alderson and Wall. They posed the question if the washback exists. Since then, the efforts of the scholars have resulted in five main models of washback. The first is the model proposed by Alderson and Wall in 1993. They argued that the test influences learning content, learning rate, learning strategies, and attitudes towards learning. Hughes described the second model in 1993. He emphasized that the test affects three elements of the learning process. These include processes, participants, and products of learning and teaching. Bailey proposed the third model in 1996. The model emphasized that students are stakeholders in the learning process. They are indirectly influenced by other participants' processes. Shih proposed the fourth model in 2007. The model classified the influences of test into three sets of factors. These include intrinsic factors, extrinsic factors, and test factors. The three factors have mutual influences on each other and an overall influence on the learning of students. Green proposed the fifth model in 2007. The model emphasized on washback directions of variability, intensity, and test shakes. It was argued that a positive washback can be achieved through an overlap between the focal construct and test characteristics (Xiao, 2014).

The concept of washback is associated with the idea of whether the successfulness of candidates in a test is an indicator of the development of life skills that are needed for everyday communication. A standardized test such as International English Language Testing System (IELTS) claims to be a test of proficiency. However, its influence on learning and teaching is dependent on the development of life skills required for social communication.

Based on the overview and background of the topic, the researcher has formulated the following research questions for the study. The review of the literature also assisted the literature in the formulation of research questions of the study.

1. What are the effects of washback in learning of language?
2. What are the effects of washback in teaching of language?



MJAL 7:2 Summer 2015

ISSN 0974-8741

Effectiveness of washback in language testing: analyzing the validity and reliability by Mohammad Altowaim

3. How can the validity of washback be analyzed?
4. How can the reliability of washback be analyzed?

2. Materials and methods

2.1 Participants

The participants of the research were English as a Foreign Language (EFL) students. Sample consisted of 50 learners. Twenty-five learners belonged to the institutes that were teaching measurement-driven or test-based English. These institutes would provide the certification of completion only when the student has successfully cleared the test. The test was the only measure of the English learning. The other twenty-five learners belonged to the institutes that would provide a certificate of completion only based on the attendance. If the students have attended at least 75 percent classes, they will be provided certificates of the completion of the course. The age of the students ranged from 13 years to 18 years. They had Arabic, Chinese, and French as their first languages.

2.2 Instruments

The instrument of data collection was a face-to-face interview with the students. An expert in English language asked questions about the language from each participant. The expert has been made aware of the questions that are usually asked in the tests of English language. He intentionally asked those questions as well that were not included in any of these tests. The purpose was to analyze if the knowledge of the students who attended test-based classes is limited to the test questions only, or they have an overall understanding of the concepts and constructs. The questions were related to grammar, vocabulary, sentence comprehension, sentence completion, and identification of errors. The language expert gave the score to each participant from 1 to 100.

2.3 Data Analysis

In order to analyze the data, the researcher used inferential statistics. The following hypothesis was put to the statistical test:



Effectiveness of washback in language testing: analyzing the validity and reliability by Mohammad Altowaim

H₀: There is no statistically significant difference in the average scores of students who attended test-based classes and who attended general classes.

H_a: There is statistically significant difference in the average scores of students who attended test-based classes and who attended general classes.

To test the hypothesis, independent-samples t-test was used. The test compares the means between two groups on the same continuous, dependent variable. In this study, the independent variable consisted of two categorical, independent groups. The first group comprised of students who attended test based classes. The second group comprised of students who attended general classes. If the group means are different, then the value of Sig. (2-tailed) in the Independent Samples Test table should be less than 0.05 for the confidence interval percentage of 95 percent. The researcher entered the data in Statistical Package for the Social Sciences (SPSS) and obtained the tables of Group Statistics and Independent Samples Test.

3. Results and discussion

The results of the participants are as follows:

Table 2 the result of the participants

Participant #	Class	Interview Score
1	Test Based	70
2	Test Based	60
3	Test Based	75
4	Test Based	90
5	Test Based	80
6	Test Based	60
7	Test Based	79
8	Test Based	74
9	Test Based	89
10	Test Based	65
11	Test Based	56



MJAL 7:2 Summer 2015

ISSN 0974-8741

Effectiveness of washback in language testing: analyzing the validity and reliability by Mohammad Altowaim

Participant #	Class	Interview Score
12	Test Based	82
13	Test Based	74
14	Test Based	61
15	Test Based	90
16	Test Based	63
17	Test Based	76
18	Test Based	87
19	Test Based	60
20	Test Based	50
21	Test Based	41
22	Test Based	78
23	Test Based	64
24	Test Based	81
25	Test Based	69
26	General	89
27	General	95
28	General	79
29	General	69
30	General	76
31	General	87
32	General	96
33	General	82
34	General	69
35	General	77
36	General	68
37	General	86
38	General	89



MJAL 7:2 Summer 2015

ISSN 0974-8741

Effectiveness of washback in language testing: analyzing the validity and reliability by Mohammad Altowaim

Participant #	Class	Interview Score
39	General	92
40	General	76
41	General	73
42	General	63
43	General	93
44	General	84
45	General	86
46	General	76
47	General	79
48	General	71
49	General	86
50	General	92

The independent samples t-test was run in SPSS. It produced the following two tables

Table 3 group statistics

	Type of Class	N	Mean	Std. Deviation	Std. Error Mean
Score in the Interview	Test Based	25	70.96	12.844	2.569
	General	25	81.32	9.348	1.870

The sample size (N) for both test-based and general classes were 25. The mean interview score for test-based class students was 70.96, and general class students had a mean score of 81.32. The standard deviation of the test-based data was 12.844, and general data was 9.348.



Effectiveness of washback in language testing: analyzing the validity and reliability by Mohammad Altowaim

Table 4 independent samples test

		Score in the Interview	
		Equal variances assumed	Equal variances not assumed
Levene's Test for Equality of Variances	F	2.483	
	Sig.	.122	
t-test for Equality of Means	t	-3.261	-3.261
	df	48	43.857
	Sig. (2-tailed)	.002	.002
	Mean Difference	-10.360	-10.360
	Std. Error Difference	3.177	3.177
	95% Confidence Interval of the Difference		
	Lower	-16.748	-16.764
	Upper	-3.972	-3.956

From the table, it is evident that group means are significantly different because the value in the Sig. (2-tailed) row is 0.002, which is less than 0.05. The Group Statistics table shows that people who attended test-based classes had lower interview scores than those who attended general classes. Hence, this study found that students of test-based classes had statistically significant lower scores (70.96 ± 12.844) compared to students of general classes (81.32 ± 9.348), $t(48) = -3.261$, $p = 0.002$.

The findings of the study indicate the validity and reliability of washback in language testing. When the students are made to realize that they will have to appear in a test, they tend to focus everything within the parameters of the marks of the test. The content that could have been asked in the test becomes the most important. The remaining content is ignored and



MJAL 7:2 Summer 2015

ISSN 0974-8741

Effectiveness of washback in language testing: analyzing the validity and reliability by Mohammad Altowaim

considered unimportant. It was reflected in the inability of the students who attended test-based classes. They could not answer those questions that are not asked in tests but part of the curriculum. The students of general classes, however, tend to have a holistic approach towards the course. They give equal attention and focus on complete content because there is no test to classify the content as relevant or irrelevant. Also, the focus of the general classes' students was found to be enhanced due to the absence of fear factor. The test generates a feeling of fear in students that in order to get a certificate of completion, they must be successful in the test. Else, their whole time and investment will be wasted. On the other hand, students of general classes attend the classes with a peace of mind with no fear of test. They just need to attend the classes to get the certificate of completion. The washback effect has a negative effect on the learning of students as found in this study. The effect is reliable and valid measure in calculating the effectiveness of learning and teaching.

The washback effect also influences the teaching methodologies of the instructors. Instructors tend to emphasize those aspects of the curriculum that might be asked in the test. Hence, the test guides the teaching methodology. Also, the good results of the class reflect well on the performance of the teacher. Hence, the teacher becomes interested in improving the scores of the students in the language tests. In the process, the totality of the contents of curriculum is hidden, and the teacher loses sight of the holistic nature of the course and the importance of all the themes and chapters for the lives of the students.

4. Conclusion

Washback refers to the effect of testing on learning and teaching. The concepts of validity and reliability are also applied in the context of the effectiveness of washback in language testing. The validity and reliability are measured by the extent to which the test has a positive influence on teaching. The findings of this study indicate that washback significantly affects the learning and teaching methodologies. The test divides the curriculum into the portions of relevant/irrelevant and important/unimportant. Also, it creates a fear factor in students during the process of learning. The teachers are also inclined to emphasize on contents relevant to the



MJAL 7:2 Summer 2015

ISSN 0974-8741

Effectiveness of washback in language testing: analyzing the validity and reliability by Mohammad Altowaim

test. They get interested and attracted in improving the grades and scores of the students. The broader vision of imparting knowledge and quality education is lost in the process.

REFERENCES

- Laborda, J. G., Litzler, M. F., Such, J. G., Bakieva, M., & Juan, N. O. D. (2012). Washback Factors and Inference Predictors in the Spanish University Examination: The OPENPAU Project (FFI2011-22442). *Procedia-Social and Behavioral Sciences*, 46, 5914-5918.
- Liu, F., & Stapleton, P. (2014). Counterargumentation and the cultivation of critical thinking in argumentative writing: Investigating washback from a high-stakes test. *System*, 45, 117-128.
- Safa, M. A., & Goodarzi, S. (2014). The Washback Effects of Task-based Assessment on the Iranian EFL Learners' Grammar Development. *Procedia-Social and Behavioral Sciences*, 98, 90-99.
- Tayeb, Y. A., Aziz, M. S. A., Ismail, K., & Khan, A. B. M. A. (2014). The Washback Effect of the General Secondary English Examination (GSEE) on Teaching and Learning. *GEMA Online Journal of Language Studies*, 14(3).
- Tomlinson, B. (2013). *Applied linguistics and materials development*. A&C Black. Retrieved from <https://books.google.com/books?isbn=1441109439>
- Xiao, W. (2014). The Intensity and Direction of CET Washback on Chinese College Students' Test-taking Strategy Use. *Theory and Practice in Language Studies*, 4(6), 1171-1177.