

Exploring The Washback Concept In Language Testing

Oleh : Tiara Dyah Andriani

Corresponding email : tiaradyah04@gmail.com

ABSTRACT

This paper is a literature review which aims to dig more exploration about washback concept in language testing. The definition of washback in language teaching is also explored. Some definitions of washback are provided from some scholars and the construct of it are also provided. This paper also discusses about the mechanisms of washback which is related with the hypotheses of washback explained by the scholars. Furthermore, how to create positive washback and the factors supporting positive washback are also discussed in this paper. To date, the researcher provides a review about measuring washback in language teaching by digging the explanation from the researcher and constructing her argument about it. In the last part of this paper, she provides sample studies about washback in language teaching from some previous research done by the experts in this field.

Keywords: *washback, language teaching, concept*

INTRODUCTION

Washback or backwash is defined as the impact of a test has on the teaching and learning. Some researchers argue that washback is the key concern for the teachers. It also can be used as the stepping-stone for the teachers to apply the best teaching methodology in the classroom. It is better for the teacher to know further about washback of the tests in order they can create a good assessment. This paper is a literature review which explores answering some problem statements, such as what is the definition of washback, what is the

mechanism of washback, how to create positive washback, how to measure washback, and what are the related research about washback in language teaching. Furthermore, this paper aims to dig more review about washback taken from some previous research in this field.

The Definition of Washback

The common assumption about tests is the teachers believe that they will be influenced by the knowledge that their students are planning to take a certain test and they will adapt their teaching methodology and lesson content to reflect the test's demands.

The term that has been used to refer to the way a test affects teaching materials and classroom management called washback. The term of washback has been used widely today within the applied linguistics and language testing. The common definition about washback is the influence of testing on teaching and learning (Alderson & Wall, 1993; Gates, 1995; Cheng & Curtis, 2004).

Shohamy et al (1996, p. 298) defines washback as “the connections between testing and learning”. He argues that there is a relationship between testing and learning that can be explored. In addition to that, Buck (1998, p. 17) describes washback as the influence of the test in the classroom. It is believed that there is a tendency for the teachers and students to conduct their classroom activities to the demands of the test, for instance if the test is very important to the future of the students and pass rates are used as a measure of teacher success.

Similarly, Hughes (1989, p. 1) defines washback simply as the effect of testing on teaching and learning. Meanwhile, Messick (1996, p. 241)

argues that washback can have either positive or negative effects. He describes it as “the extent to which the introduction and use of a test influences language teachers and learners to do things they would not otherwise do that promote or inhibit language learning”.

From those definitions of washback, it can be defined that washback is the effect of testing on teaching and learning that has tendency for both the teachers and students to conduct their classroom activities to the demands of the test, in addition to that washback can have either positive or negative effects.

Mechanisms of washback

Hughes (1993, p. 2) discusses the mechanisms of washback. He states that to distinguish between participants, process, and product in teaching and learning might be helpful to recognize that all three may be affected by the nature of a test. He defines that participants include the students, classroom teachers, administrators, materials developers, and publishers. The processes here include materials development, syllabus design, changes

in teaching methodology, the use of learning and / or test-taking strategies, etc. According to him, product refers to “what is learned” (facts, skills, etc) and the quality of learning, such as the fluency.

Alderson and Wall (1993, p. 120-121) propose different approach about how washback works. They propose 15 possible hypotheses regarding washback, such as:

- a. A test will influence teaching

This is the most general hypotheses of washback. This is based on the assumption that teaching and learning are related. Thus, the second hypothesis follows the implication from the first one.

- b. A test will influence learning

To separate the content of teaching from its methodology, we need to distinguish the influence of a test on the content of the teaching from its influence on the methodology. Thus:

- c. A test will influence what teachers teach

- d. A test will influence how teachers teach

- e. A test will influence what learners learn

In order to consider how quickly and in what order teachers teach and learners learn, so:

- f. A test will influence the rate and sequence of teaching

- g. A test will influence the rate and sequence of learning

As a teacher, we need to consider both the quality and the quantity of teaching and learning, so:

- h. A test will influence the degree and depth of teaching

- i. A test will influence the degree and depth of learning

If washback relates to attitude and behaviours, so:

- j. A test will influence attitudes to the content, method, etc of teaching and learning.

From the previous hypotheses, there is no consideration to the nature of the test, furthermore to the uses of to the which scores will be put.

- k. Tests that have important consequences will have washback, and conversely.
- l. Tests that do not have important consequences will have no washback.

Then,

- m. Tests will have washback on all learners and teachers.
- n. Tests will have washback effects for some learners and some teachers, but not for others.

Creating Positive Washback

It has been clearly discussed above that the washback can be either positive or negative. The positive or negative washback appears depending on the way the teachers distributed their tests. It can be related to the validity of the instruments, the situations in the classroom, the physical condition of the students or any other factors which affect to the result of the test. Messick in Ozmen (2011, p.217) states that a positive washback effect can be created if the tasks and activities used in teaching language are included as test items in the examination. It means that

the students should not be only tested on what they were taught, but also they should be tested in the way or they were taught.

The way means the methodology used by the teacher during the teaching and learning process before the students are tested. Washback from tests can involve individual teachers and students as well as whole classes and programs. However, most of the teachers want to create a positive washback in their assessment. Hughes (1994) advised that at least five conditions have to be met before all of the possible washback effects can occur:

- a. Success on the test must be important to the learners,
- b. Teachers must want their learners to succeed,
- c. Participants must be familiar with the test *‘and understand the implications of its nature and content’*,
- d. Participants must have the expertise which is demanded by the test (including teaching methods, syllabus design and materials writing expertise), and

e. The necessary resources for successful test preparation must be available (ibid: 2-3)

Moreover, several research have been done in order to know the steps and the factors of creating positive washback. According to Bailey (2014, p.268) there are a number of factors to create positive washback. These include the incorporation of 1) language learning goals; 2) authenticity; 3) learner autonomy and self-assessment; and 4) detailed score reporting, each of which will be explained briefly below.

a) Language learning goals

As we know that the implementation of the tests in teaching and learning aims to determine the results of student learning outcomes. In addition, it is also related to whether the student is able to achieve the learning goals or not. The test will create positive washback to the extent if it promotes the accomplishment of educational goals held by learners and/or programme personnel. Buck (1988, p.17) defines the

relationship between tests and educational goals.

He states that most educators would probably agree that the content of classroom instruction should be decided on the basis of clearly understood educational goals, and examinations should try to make sure whether these goals have been achieved. As the result, it forces students and teachers to focus on these goals, and the washback effect on the classroom is achieved.

b) Building in authenticity

Actually, a test must be contextual and can be associated to the real life in order the students' learning outcomes are meaningful. The students can also apply the solution which they use to answer the question in the test to their real life. The contextual in the test related to the authenticity. Doye (1991, p.104) defines authenticity as the term which is used in language testing. He states that absolute congruence would exist when the tasks in the test situation and in the corresponding real-life

situation would actually be identical.

Moreover, the test situation and the tasks in it are called authentic. An authentic test is therefore one that reproduces a real-life situation in order to examine the student's ability to cope with it. It is also stated by Morrow (1991, p.112) the most striking implication of adopting the perspective on the design of a language test is the overriding importance of authenticity both of text (as input) and of task (in processing this input). It means that the input and output must be equally required.

To achieve the equal requirements, the test should be contextual to the real life authentic so that the students will get positive washback of the test. Messick (1996, p.241) discusses the connection between authenticity and washback. He points out that tests having positive washback are likely to be criterion samples. They should be authentic and direct samples of the

communicative behaviour of listening, speaking, reading and writing of the language being learnt.

Ideally, the move from learning exercises to test exercises should be seamless. Messick (1996, p. 252) also notes that there are two possible threats to validity which can affect a test's authenticity. The first is construct under-representation (when the assessment mechanism does not include key features of the construct being measured) and the other is construct-irrelevant variance (when the assessment includes 'excess reliable variance that is irrelevant to the interpreted construct'). Both threats, are present in any test to some degree. They are tied to washback in that 'if one is concerned with fostering positive washback and reducing negative washback, one should concentrate first on minimizing construct under-representation and construct-irrelevant difficulty in the assessment.

c) Learner autonomy and self-assessment

Learner autonomy and self-assessment ensure the positive washback. According to Bailey (2014, p.270), learner autonomy refers to the philosophy that learners should have a large amount to say about what, how and how fast they learn. It incorporates principles of choice, intrinsic motivation, attention focus and personal evaluation. In addition, it is absolutely believed that the teaching activity is related to the teaching test.

The way students learn in the classroom will affect the results of learning outcomes. Autonomous learners develop their own internal values with regard to judging progress on the material and/or skills to be learnt. It is supported by Alderson and Wall's that a test will influence the degree and depth of learning' (1993, p.120). Self-assessment can be related to beneficial washback because it helps learners develop internal criteria for progress and success,

and thus develop learner autonomy. Furthermore, von Elek (1985, p.60) explains the direct relationship of self-assessment to autonomous learning and positive washback, as follows:

- a. It enables learners to assume greater responsibility for the assessment of their proficiency and their progress;
- b. It enables them to diagnose their weak areas and to get a realistic view of their overall ability and their skills profile;
- c. It enables them to see their present proficiency in relation to the level they wish to attain;
- d. It helps them to become more motivated and goal oriented.

Thus self-assessment and learner autonomy are linked to each other and to washback, because developing internal criteria for success is one of the key characteristics of autonomous learning.

d) Score reporting

Standard test can provide positive washback both to the learners themselves and to

programme representatives. To conduct a standard test the teachers should minimally entail more detailed score reports which are typically available with wide-scale proficiency tests. According to Shohamy (1992, p.515) to create positive washback, assessment information must be 'detailed, innovative, relevant and diagnostic' and that it must 'address a variety of dimensions rather than being collapsed into one general score' (1992: 515).

Spolsky (1990: 12) explains about the relationship between the use of detailed score reports and washback. He states that there is a natural tendency on the part of those who use test results to take shortcuts, there is a moral responsibility on testers to see that results are not just accurate but do not lend themselves to too quick interpretation.'

The issues of authenticity, autonomy and self-assessment, and detailed score reporting are widely held in the available literature to create positive washback. And yet, a further empirical research needs to

be conducted to substantiate these claims.

Measuring Washback

Generally, it is difficult to see how washback could be measured. The term of measuring washback is identically used in quantitative, experimental approaches to research. The questionnaires as the data instruments might be used to get the correct measurement of washback. The difficulty in measuring washback occurs because of the apparent inseparability of washback from most other features of teaching and learning.

This problem runs counter to a basic tenet of measurement in the experimental tradition, which entails identifying, operationally defining and controlling the variables that impinge upon the desired measurements. Messick has noted (1994,p.16) that interpretation of such teaching and learning consequences as reflective of test validity (or invalidity) assumes that all other aspects of the educational system are working well or are controlled. Thus, in practice the issue is not just the systematic validity of the

tests but rather the validity of the system as a whole for improving teaching and learning. It appears that, except in the most artificial, laboratory-like conditions of experimental research, washback cannot be entirely separated from other variables that influence teaching and learning.

There are some problems in measuring washback, as stated by Bailey (2014, p.272). In examining washback the teachers are considering the interface of measurement procedures with the real world using experimental validity continuum. It sets in naturally occurring settings (language programmes) with a nonrandom sample of subjects. The sample of subjects refers to the language learners who choose to enrol and in some instances to take the test. The teachers also use face-validity to make sure that the test is considered as a good test. They check the questions provided in the test with the syllabus and blueprint which congruent to the curriculum.

In contrary, Alderson and Wall (1993, p.127-128) state that trying to develop a true experimental design to

assess washback would be both futile and unhelpful. They offer 'a series of proposals for research'. Among their methodological suggestions, there is a heavy emphasis on classroom observation and triangulation. In investigations of language testing washback, triangulation would minimally include teachers' and students' perceptions of what have called 'washback to the programme' and 'washback to the learner'. Moreover, it can be assumed that to measure the washback, it doesn't only need statistics in analyzing data. The qualitative one such as the triangulation to check the test of language teaching is also required.

Sample Studies of Washback

Some studies have been done to investigate the phenomenon of washback and explore how it affects the various stakeholders. Messick (1996, p.242) states that the most complete designs for researching washback include both observing classes and asking the participants about their views and experiences (either through interviews or written questionnaires) to

determine whether teaching and methodology which was teacher-learning are 'evidentially linked to the centered remained unchanged. introduction and use of the test. Cheng Cheng argued that the reason why (1999, p.268-269) has investigated the teachers did not change teaching washback effects of the 1994 Revised methodology was that the examination Hong Kong Certificate of Education did not influence teachers' beliefs about Exam in English (HKCEE) in the teaching and learning. context of secondary school examinations in Hongkong. Cheng has investigated the impact both at the macro level on the different stakeholders like the Examination Authority, textbook publishers, and the tertiary institutions; and at the micro-level on the teachers' attitudes, teaching content and classroom interactions. She undertook classroom observations of 45 lessons of 12 teachers, used questionnaires for 550 teachers and 1700 students, and conducted interviews of an unspecified number of teachers. Her data revealed a range of attitudes and behavioral changes.

The examination was supposed to encourage task-based learning. The result of the investigation was the teachers felt positively about the examination that enabled students to use English more practically and authentically. However, the teaching practice and that little positive washback was felt on students' language use. Munoz and Alvarez (2009, p.1) reports the results of a research study to determine the washback effect of an

oral assessment system on some areas of the teaching and learning of English as a Foreign Language (EFL). The research combined quantitative and qualitative research methods within a comparative study between an experimental group and a comparison group. They took fourteen EFL teachers and 110 college students as the participants in the study. Data were collected by means of teacher and student surveys, class observations, and external evaluations of students' oral performance.

The data were analyzed using descriptive statistics for qualitative information and inferential statistics to compare the mean scores of the two groups by One Way Anova. The results showed that there are positive washback in some of the areas examined. The implications for the classroom are that constant guidance and support over time are essential in order to help teachers use the system appropriately and therefore create positive washback.

From the studies above, it can be assumed that merely instituting examinations is unlikely to bring about

significant change. For meaningful educational change a host of elements beyond the examination itself needs to be considered. To support the statement, there is a study on Sri Lanka investigated by Wall (1996) as cited in Munoz and Alvarez (2009, p.4). He found that there were several different factors which might have prevented the examination from providing positive effects: Teachers' lack of understanding of the exam, resistance to change, and exam content. She also refers to other factors such as gap between test designers and teachers, and lack of well trained teachers.

There are also some studies which reveal about the negative washback. Sanel and Tutunis (2011, p. 47) conducted a research consisted of 42 students at the English Preparatory School of Istanbul Aydın University. Of all the participants, 19 were male and 23 were female students. They all studied English during their high school education. The study took place for 2 weeks. 2 classes (21 students in each class) were chosen randomly as a control and an experimental group.

A questionnaire was given to both

groups to see their beliefs and attitudes towards writing exams. Three main results obtained from this study. The first one is that EFL learners experience of testing anxiety is aroused by some main factors, such as lack of word knowledge; lack of grammar and syntax knowledge; fear of failure; fear of negative evaluation; anxiety and time limit. Secondly, the sources of language anxiety consist of fear of failing class; fear of performing badly with regard to classmates; fear of making mistakes; fear of using wrong expressions. Thirdly, language learners feel more comfortable and unstressed when teachers do not let them know that it is an assessment. In other words, EFL learners will not be anxious and excited if teachers do not let them know that it is an assessment.

Another study is conducted by Ozmen in 2011. Ozmen (2011, p.211) presented a study investigating the washback effect of Inter-university Foreign Language Examination (ILE) on candidate academics in Turkey. The present study follows twelve candidate academics during their preparation for ILE in qualitative design. The results

revealed that ILE exerts a negative impact on candidate academics.

The results also show the exam does not only have a micro level effect, which can be defined as the effect on an individual or a small group of individuals, but also a macro effect on a relatively populated group of individuals studying or working at university contexts. Participants also reported inhibition and avoidance problems, which can be regarded as a very critical problems for academics whose primary language is mostly English in both studying, publishing and presenting. In this study, the reason why ILE exerts the harmful and negative washback effects on the candidate academics is attributed to the content and style of the test itself, which is a multiple-choice test claiming to assess grammar, vocabulary and reading-skills knowledge in the L2 – English in the context and scope of this study. The unanimous reactions of all participating candidate academics of this study were that they need to develop more than what ILE assesses, and that ILE represents an obstacle in front of them.

To sum up, the results of the washback of the test are varied depends on the way the teachers distributed the tests. In measuring and deciding the washback, the researchers used quantitative terms such as questionnaire and checked with observations to make the results gained the trustworthiness. Finally, it will be possible if there are numerous opportunities for future research on language testing washback.

CONCLUSION

Washback is the effect of testing on teaching and learning that has tendency for both the teachers and students to conduct their classroom activities to the demands of the test. Washback is actually complex and elusive; in some studies showed that there can be washback from exams onto a variety of teaching and learning areas. There are two model of washbacks namely positive washback and negative washback. The positive or negative washback appears depends on the way the teachers distributed their tests. It can be related to the validity of the instruments, the situations in the classroom, the physical condition of the students or any other factors which

affect to the result of the test. Some studies have been conducted to investigate the washback of the tests. The discussion of studies focus on the washback of oral assessment, written assessment or the evaluation of curriculum in certain countries. Moreover, the future research on language testing can be possibly done in order to reveal more facts about washback.

REFERENCES

- Aftab, Asma. Qureeshi, Saben. & William, Isabel. (2014). Investigating the Washback Effect of the Pakistani Intermediate English Examination. *International Journal of English and Literature*. Vol 5(7) pp. 149-154.
- Alderson, J.C & D. Wall. (1993). Does washback exist?. *Applied Linguistics*, 14(2), 115-129.
- Buck, G. (1988). Testing listening comprehension in Japanese university entrance examinations. *JALT Journal*, 10, 15-42.
- Cheng, L. 1999. 'Changing assessment: Washback on teacher perceptions and actions'. *Teaching and Teacher Education* 15: 253-271
- Cheng, L. & Curtis, A. (2004). *Washback or Backwash: A review of the impact of testing on*

- teaching and learning*. In Cheng, L., Y., Watanabe, & A. Curtis.
- Doye, P. (1991). Authenticity in foreign language testing. In Anivan, S., editor, *Current developments in language testing. Anthology Series 25*, Singapore: Regional Language Centre, 103-10.
- Gates, S. (1995). *Exploiting Washback from standardized tests*. In Brown, J.D & S.O. Yamashita (Eds.).
- Hughes, A. (1989). *Testing for language teachers*. Cambridge: Cambridge University Press.
- Hughes, A. (1993). *Backwash and TOEFL 2000*. Unpublished manuscript: University of Reading.
- Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13(3), 241-256.
- Morrow, K. (1991). Evaluating communicative tests. In Anivan, S., editor, *Current developments in language testing. Anthology Series 25*, Singapore: Regional Language Centre, 111-18.
- Muñoz, Ana P. & Alvarez, Marta E. (2009) Washback of an oral assessment system in the EFL classroom. *Language Testing* 27(1) 1-1.
- Özmen, Kemal Sinan. (2011). Washback Effects of The Inter-University Foreign Language Examination on Foreign Language Competences of Candidate Academics. *Novitas-ROYAL (Research on Youth and Language)*, 2011, 5(2), 215-228.
- Şenel, Esmâ & Tütüniş, Birsen. (2011). *The Washback Effect Of Testing On Students' Learning In Efl Writing Classes*. Dil Dergisi • Sayı: 153 • Temmuz-Ağustos-Eylül 2011
- Shohamy, E., S., Donitsa-Schmidt & I., Ferman. (1996). Test impact revisited: Washback effect over time. *Language Testing*, 13(3), 298-217.
- Spolsky, B. 1990: Social aspects of individual assessment. In de Jong, J.H.A.L. and Stevenson, D.K., editors, *Individualizing the assessment of language abilities*, Clevedon: Multilingual Matters, 3-15.
- von Elek, T. 1985: A test of Swedish as a second language: an experiment in self-assessment. In Lee, Y.P., Fok, C.Y.Y., Lord, R. and Low, G., editors, *New dimensions in language testing*, Oxford: Pergamon Press, 47-58