

Washback in Language Testing: Review of Related Literature First

Neda Beikmahdavi

Department of ELT, Ayatollah Amoli Branch, Islamic Azad University, Amol, Mazandaran, Iran

Corresponding author's email: beikmahdavi@yahoo.com

ABSTRACT

Assessment is a significant section in the process of language teaching and learning. One important part of language testing is the outcomes of the tests. The positive or negative outcomes of the test is associated with washback. In the present study, the main purpose was to review the conceptual underpinnings of washback. Washback is a useful means to consider the results of the test effects on teacher and learner behavior in the classroom. It is divided into two main categories: positive and negative washback.

KEYWORDS: washback, language teaching, language testing, learner behavior

INTRODUCTION

During the history of language teaching and testing, language test can be a crucial tool and it is considerably applied. In addition, language tests are considered as a valuable means for providing information that is related to various concerns in the process of language teaching and learning. They can be used as a method for supplying systematic feedback for both teachers and learners. Also, the educator can observe how well or badly the learners have performances and checking any differences between expectations and real performance. Besides, the learners can know how much achievement and progress they are carrying out in acquiring the language.

Tests in language can also be a useful instrument for assessing instructional materials and tasks and their link to the educational objectives. In ideal form, the objectives of the tests in language should be obvious for learners, in order that they require not spend time supposing what the educators means. If the tests are relevant to the learners' needs, they are possibly trying to engage more dynamically in the language learning process.

Another dimension of language tests concerns is the outcomes which are taken by a teacher's insights and inferences (Bailey, 1996). The effectiveness of such interferences can be obvious when they provide feedback to be applied in making the program of teaching more efficient and when they provide information to the kinds of materials and tasks learners require. These insights can be based on the teachers' appropriate decisions to the teaching operation.

For achieving its goals, a test in language must encounter the needs of some necessary test-qualities, such as practicality, validity and reliability. In other contexts, it should be interactive and authentic and also should have an effect on all parties.

In addition, a basic assumption behind the numerous utilities of language tests is that a language test reflects an effect on both learning and teaching. In the literature of the assessment, there is an obvious consensus that the influence of a test on learners and teachers is termed washback. Lot of experts believe that the concept of washback is directly related to the principles that some influences of a language test may be efficient, and some other effects may be harmful. A positive washback is used to encourage the process of learning; in a contrary manner, a negative washback usually inhibits the achievements of the educational purposes held by students, as Alderson(2004). In the present study, the main purpose is to discuss the theoretical underpinnings of washback. In particular, different definitions, origins and categories and functions of washback are taken into account.

DEFINITIONS OF WASHBACK

In language testing, different researchers presented different washback differently. These definitions can be simple and complex. as Wall (1997), the term washback is restricted to "test effects on teacher and learner behavior in the classroom whereas impact refers to wider test effects such as their influence on teaching materials and educational systems" (p. 100). as Messick (1996), "washback refers to the extent to which the introduction and use of a test influences language teachers and learners to do things that they would not otherwise do that promote or inhibit language learning".as Bachman and Palmer (1996) presented a definition about washback and test impact in a similar fashion and viewed washback as one dimension of test impact. They also add that washback is a more

complicated phenomenon than simply the impact of a test on teaching and learning. Therefore, they recommended that the effect of a test should be assessed with reference to contextual variables like as society's values and purposes, the educational system in which the test was applied, as well as the potential results of its use. as Hawkey (2006) points out the concepts washback and impact are usually taken into account in terms of their logical location, definition scope, positive and negative implications, intentionality, complexity, direction, intensity, relationship with validity and validation, relationship with testing critical view, and the role in washback/impact models.

According to the above mentioned definition, it can be concluded that washback is referred to the effects of test on the learners and teachers' behaviors in the classroom context.

THE ORIGIN OF WASHBACK

The origination of washback traced back to beginning of the 1990s. Before, applied linguistics applied various terms to refer to the idea of examination effect. Some of these terms can be: "test impact" as Bachman and Palmer (1996), "systematic validity" as Messick (1989); "measurement-driven instruction" as Popham (1987), "curriculum alignment" as Shepard (1990), and also "backwash" as Biggs (1993). Among different terms, two terms are applied dominantly and they are as: washback and backwash. Different scholars use interchangeably together. To approve this interpretation of the two concepts, as Hughes (1993) states that there is an interchangeable use of the two expressions. He makes it more explicit when he points out that "where washback comes from, I don't know". That I know is that you can find backwash in dictionaries, but not washback" (p. 57).

Investigators in language testing have perceived that the emergence of this concept is the outcome of the significant reforms and advances that have participated in the language testing field at the end of the twentieth century. As Cheng (2005), the subject of examination effect was rooted in the notion of that tests are usually observed to move teaching and learning. He also claimed that this is necessary to create the matching between the test construct and what the teachers presents in instruction. As Messick (1989), the concept of washback should be in the wider realm of construct validity. He argues that this construct involves a number of dimensions about testing like as the impact of tests of language test takers and educators, the interpretations of scores by decision makers, and the planned uses of test scores. According to this view, the washback concept is associated with an inherent quality of any assessment type, particularly, when the future of test takers are influenced by examinations outcomes.

Based on the significance of Washback in the field of language teaching and testing, as T sagari (2006) recommended an artificial time framework are categorized into three distinguishing but successive steps: a) the "pre-1993" step; b) the "1993 step"; and c) the "post-1993" step. Based on the first step, he labelled the "pre-1993" step the "myth" step. It was based on the period of time when authors determined the examination influence phenomenon but no one took into account for it. as Hughes (1989), the considerable point in this era was that a few number of studies were conducted and published to the language testing community which made strong arguments of the absence of this phenomenon. In this period, most of these empirical studies were depended upon self-report data or on direct results or on test outcomes rather that direct contact with participants included. In the second step, which was markedly different from the prior phase since it was typically attributed by the publication of a seminal work paper by two great language testing investigators as Alderson & Wall (1993), who are indebted the fact they were the first who questioned the feature of examination influence. The authors recommended a set of relevant hypotheses to re-conceptualize the phenomenon. A last, the third phase, which T sagari called it as the "reality step", was basically specified to be new era where fundamental research projects on washback utilized developed models to accurately elaborate the different elements that make up this concept.

In conclusion, however relatively little has been written about the origins of washback, a great deal of information has been emerged about different notions which refer to examination influence. In general, it can be said that the study of washback origins is vital to shape the scope of further required study in this field.

TYPES OF WASHBACK

The studies in language assessment have revealed that washback often refers to a movement in a specific direction. This movement is a primary part on the use of this phenomenon to explain the relationship between teaching and testing. as Pearson (1988), "public examinations influence the attitudes, behaviors, and motivation of teachers, learners, and parents, and because examination often come at the end of the course, this influence is seen working in a backwash direction, hence, the term washback"(p. 98).

Moreover, washback is considered as bipolar concept- either positive (effective) or negative (harmful). Based on as Alderson and Wall's (1993) definition about washback who defined washback as "the extent to which a test affects language teachers and students to do things they would not necessarily otherwise do that promote or inhibit

language learning as Messik (1996). They also add that "tests can be powerful determiners, both negatively and positively, of what takes place in classroom contexts (p. 245).

Based on this explanation, there are two types of washback as negative and positive which are explained in the next section. Negative washback is considered as the negative influence of tests in the process of language teaching and learning. as Alderson and Wall (1993), negative washback is defined as "the undesirable effects on teaching and learning of a specific test. The tests may fail to reflect the learning principles and/or the course objectives to which they are supposedly related (p. 5).

Based on this definition, such tests will lead to the narrowing of content in the curriculum rather than covering a definite content what has been acquired in the classroom context. As Vernon (1956) states that educators prefer to neglect subjects in activities that are not directly associated with passing examination, and testing accordingly change the curriculum in a negative format as Cheng (2005). It can be stated that those examinations may fail to provide a relationship between the learning tenets and course objectives to which they should be linked as Cheng (2005). Another influence of negative washback is that basically it can reduce the time provided for teaching instruction, narrowing curriculum offering, and instruction modes, and potentially decrease the capacities for educators to teach content and utilize materials and methods that are in congruous with effective testing tools as Smith (1991).

as Wiseman (1961), one of the strong effect of negative washback is that a growing number of coaching classes are set up to make ready learners for tests, but what learners will acquire are test-taking skills instead of language tasks and activities as Bailey (1996). In this context, Shohamy, Donitsa-Schmidt and Ferman (2004) point out that high fear and anxiety of examination results will be increased among students and teachers. They believe that teachers sense that failure or success of their learners is reflected on them, and they usually speak of pressure to cover the materials for the test. When the learners know that one single measure of performance can specify their levels, they will less possibly a positive attitude toward learning.

On the other side of washback concept is positive side. According to Cheng and Watanabe (2004), tests can provide beneficial changes in teaching and specifically in examinations, reflecting the positive washback. As Pearson (1988) declares that good examinations can be designed and applied as effective teaching-learning tasks and activities in order to encouraging a positive teaching-learning process. As Andrews (1994) recommends innovations in the curriculum of language through modifications in language testing. For example, an oral proficiency test was introduced in the expectation that it would promote the teaching of speaking. as Davies (1985), "an innovative and creative test can efficiently result in syllabus alteration or even in a new syllabus" (p.18).

However, in educational settings, most experts argue that there is no obvious consensus among practitioners as to whether specific washback effects are positive or negative. As Hsu (2009) points out that "one justification to this struggling situation is that potentially negative or positive nature of the test can be affected by a lot of contextual factors" (p. 9). Alderson and Wall (1993) mention that the quality of washback effect might be stated to be beneficial or detrimental.

Thus, a lot of washback studies may be beneficial and they provided positive alterations in the process of teaching and learning. Unlike these positive changes, some other studies revealed the negative effects of washback studies. As Alderson and Wall (1993) found that the best way to recognize the determined effect is to study as thoroughly as probable the wide educational context in which the act of evaluating is taking part, since the main variables that usually influence this act exist within the educational system, and might prevent washback from appearing. as Cheng (2005) give a summary about this situation and mentions that "if the results of a specific test for which testing and learning are to be assessed, the educational context in which the test require to be fully understood" (p. 31). The positive or negative effect of the test can be depended upon how and where it exists and manifests itself within a specific educational context so as to comprehend the mechanism of washback. In the next sections, at first, the functions of washback and next mechanisms of washback will be taken into account.

FUNCTIONS OF WASHBACK

In the past, tests were applied based on the determined goal in teaching and learning context to provide right diagnosis of the effects of teaching and learning. However, with the changes and alterations in the testing field and how this latter is conceived, a test can also be developed to be utilized in the starting or in the middle of the teaching and learning processes so as to affect either or both processes.

Based on the functions of language tests through which washback takes place in actual teaching and learning contexts, as Wall (2005) examined a set of reviews of those tests and effects they could have on the systems they are introduced into. As Eckstein and Noah (1993) produced one of these important reviews. They provided a historical account of a lot of functions and effects of some kinds of tests which showed properly how people beyond

history have commonly regarded tests as a prominent means by which they take the desired decisions and plans for some determined objectives. For example, for the writers, the first documented use of written, public test systems took place under the Han dynasty in China about 200 B.C. The major function of these specific examinations was to choose candidates for entry into government services.

The second function of washback as Eckstein and Noah (1993) explained was that one which sought to control corruption. The instance of this function was Britain where people could get entry into higher levels of education or the carrier of strengths. A significance, direct result of this test and examination was the establishment of a lot of schools, which intended to prepare learners for these tests. Moreover, the third example of function based on Eckstein and Noah's (1993) view was to increase the competence and knowledge levels among those individuals who were entering government services or carriers. The plan was to design and develop tests which showed the target situations needs and demands, learners for those tests could have to increase skills which were relevant to the work they hoped to get in the future.

The next function was that of assigning additional places in higher education. In this level, tests were applied as tools of choosing the most able candidates for the accessible places. This kind of tests can be the same to what is referred to as placement tests in the literature of testing in the present era. In this illustration, the fifth function was to gauge and foster the teachers and schools effectiveness. As Eckstein and Noah (1993) provided an instance depicting how, in the specific period of time, the government sets up an examination system to check the schools performances by those people who support these tests through the allocation of significant funds. The amount of funds that schools received depended upon how its learners performed. Although, this system had some crucial unplanned outcomes and ultimately had forced to access the determined goals. The last function was to limit curriculum differentiation. In the nineteenth and twentieth centuries, in Britain, there was a considerable resistance to the idea of centralized education, and all the schools had the freedom to plan on their own curriculum and assessment tools.

In the modern societies, frequently, tests are utilized for responsibility within the system, and particularly for certification of achievement in education. They take shape part of the procedure for decision about the allocation of few repertoires of both an individual and systematic level. For instance, in different countries and Algeria is one of them, tests monitor the transition between school and higher education.as Hsu (2009), also test are observed as ways to improve knowledge and enhance the performance of institutions. The results of tests which are visible and ideally measurable can be reported in connection with the public can understand and can be applied to reveal that shift has or has not occurred.

The above explained functions are typical situations where these examinations were utilized to indicate effect on the final consequences to match the determined goals of those who are in authority to make and impose their policies. Some testing experts declare that this is a specific common practice in countries with centralized educational systems, where the trained programs are monitored by central agencies.

Based on the tests and examinations status in public realm, it appears that it is significant to comprehend the testing functions in relation to many aspects and scopes of teaching as mentioned in the instances discussed earlier. Based on the explained functions of tests, it can be necessary to point out the different mechanisms within different educational contexts. In the next section, the mechanisms of washback are considered.

WASHBACK MECHANISMS

In the present section, the complicated mechanisms through which washback takes place in the process of teaching and learning are explicated. as Hughes (1993), there is a trichotomy to reveal how this phenomenon works in various contexts as Bailey (1996).Bailey mentions that this special trichotomy permits educators in general education and language testing experts in particular to improve a main washback model which elaborates how the different elements that make-up this framework interact to assist the comprehending of the nature of this subject of interest.

For describing this model, as Hughes (1993) mentions that the trichotomy is shaped in three parts: a) first, the subjects who are basically the individuals such as classroom-teachers, learners, administrators, material developers, and even publishers whose perceptions and attitudes toward their works may be a test. b) In this framework, the second component is called process. Any action taken by the subjects are called the process, and they may contribute to the learning process. c) Based on Hughes' model, the third section is product. It is associated with what is acquired as the skills, facts, and other aspects and also the learning quality.

Unlike Hughes (1993) who emphasizes more on the three elements that make-up this model, as Alderson and Wall (1993), concentrate on what they referred to as "micro-aspects" of the teaching and learning process and they might be influenced by examinations. They claimed that there is little evidence supplied by empirical research to

support the idea that examinations influence on teaching. They advocated that "the concept is not well-defined, and we believe that it is significant to be more precise about what washback might be before we can seek out its nature and whether it is a natural or unavoidable outcomes of testing" (p. 117).

Ultimately, they recommend 15 hypotheses that can assist investigators to indicate realms in teaching and learning that are often influenced by washback and can stand as a foundation for further investigator and investigation. These hypotheses have revealed that there exist a strong relationship between the tests significance and the extent of washback. Alderson and Wall resulted that further studies are required and they must involve enhancing specification of washback hypothesis. They reported that investigators had to take into account to research literature in the language testing field in at least two realms: performance, motivation, and also the new innovation in the educational settings. The hypotheses were:

- A test will influence teaching;
- A test will influence learning;
- A test will influence what teachers teach; and
- A test will influence how teachers teach.
- A test will influence what learners learn; and
- A test will influence how learners learn;
- A test will influence the rate and sequence of teaching; and
- A test will influence the degree and the depth of learning;
- A test will influence the degree and the depth of teaching
- A test will influence the degree of learning.
- A test will influence attitudes towards the content and method of teaching and learning.
- Tests that have important consequences will have washback; and conversely
- Tests that do not have important consequences will have no washback.
- Tests will have washback on learners and teachers.
- Tests will have washback effects for some learners and some teachers, but not for others.

Following the first studies realized on washback hypotheses, as Wall (1995) followed up their study and emphasized the complexities in finding explanations on how examinations used effect on the teaching process. Accordingly, Wall suggested that the research areas that are observed to be related to washback should involve (a) the noting of detailed framework studies to determine fundamental features in the target system and the environment, involving an analysis of the current testing practices as Bailey (1996).

In addition, in the same perspective of washback mechanisms as an event of change in teaching and learning, as Smith (1991) sought out an ELT project and worked on to construct a corresponding model of variables included with the purpose to introduce the desired shift in the teaching and learning processes. Smith's model includes five elements of change: a) the target system; b) the management system; c) the innovation itself; d) the accessible resources; and e) the environment that change is assumed to happen. as Hsu (2009), as Markee (1997) had the same belief how change might take place on larger subjects like as curricular through following phases, which are to design, to carry out, and finally to maintain. In this regard, as Markee (1997) recommended a framework that was associated with the composed questions that were asked by Cooper (1989) and they were referred to: who (subjects), what (product), where (content), when (the time, duration), why (the rationale and reasoning), and how (various approaches in managing the washback effect).

In conclusion, the reviews of different studies on the washback mechanisms have confirmed the significant correlation between the design of given examinations and their positive or negative effect and power on teaching and learning. Although, it is important to note that consequences of these studies, even if they have contributed in advancing research into the washback realm in the language testing field.

WASHBACK AND VALIDITY

A basic characteristic of washback that advantages particular consideration is its relationship to test validity. The meticulous attribute of the relationship between washback and validity has been argued. On the one hand there are authors e.g. Morrow (1986) who advocate Messick's views (1996) that the test impact on teaching and learning is considered as an important dimension of its validity.

Particularly, according to as Messick (1996), washback is located within the theoretical concepts of consequential validity in which the social consequences of testing are seen as part of a broader, unified concept of test validity: In the context of unified validity, evidence of washback is an instance of the consequential aspect of

construct validity, which is only one of six important aspects or forms of evidence contributing to the validity of language test interpretation and use (p. 254-255).

In a contrary manner, the opposite view's advocators claim that "validity is not a property of the test or assessment as such, but rather the meaning of the test scores" as Ferman (2004) and, thus, have difficulty in seeing a direct connection between washback and validity. They claim that how test scores are interpreted seems quite different and actually quite removed, for example, from how teachers teach or how students acquire before an examination happen.

CONCLUSION

This study presented the conceptual framework of the washback effect in the testing field. Washback is defined as the test effects on teacher and learner behavior in the classroom whereas impact refers to wider test effects such as their influence on teaching materials and educational systems. The origination of this concept traces back to the beginning of the 1990s. The other names of this concepts are "test impact", "systematic validity", "measurement-driven instruction", "curriculum alignment", and also "backwash". Washback is considered as bipolar concept—either positive (effective) or negative (harmful). Tests can provide beneficial changes in teaching and specifically in examinations, reflecting the positive washback. Negative washback is considered as the negative influence of tests in the process of language teaching and learning. Washback includes some functions as choosing candidates for entry into government services, controlling corruption, assigning additional places in higher education and so on. Different scholars elaborated the mechanism of washback. Washback has a direct relationship with validity. Nevertheless, the question of impact of tests on teaching and learning either in positive or negative ways still raised and unanswered, and hence needs to be continuously and thoroughly explored in further studies on testing. In particular this review of the literature on washback has revealed that actually a large portion of those studies on this phenomenon from different perspectives and multiple levels are available.

Conflict of interest

The authors declare no conflict of interest.

REFERENCES

- Alderson, J. C. (2004). Foreword. In L. Cheng, Y. Watanabe & A. Curtis (Eds.), *Washback in language testing: Research context and methods* (pp. ix-xii). Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.
- Andrews, S. (1994). The washback effect of examinations: Its impact upon curriculum innovation in English language teaching. *Curriculum Forum*, 4(1), 44-58.
- Bachman, L. F., & Palmer, A. (1996). *Language testing in practice: Designing and Developing Useful Language Tests*. Oxford: OUP.
- Bailey, K. M. (1996). Working for washback: A review of the washback concept in language testing. *Language Testing*, 13 (3), 257-279.
- Biggs, J.B. (1993). Assumptions underlying new approaches to assessment, *Curriculum Forum*, 4, 1.
- Cheng, L. (2005). *Changing language teaching through language testing: A Washback study*. Cambridge: CUP.
- Cooper, H. (1989). *Homework*. White Plains, NY: Longman.
- Davies, A. (1985). Demands of being professional in language testing. *Language Testing*, 14(3), 328-339.
- Eckstein, M. A., & Noah, H. J. (Eds.). (1993). *Examinations: Comparative and international studies*. Oxford: Pergamon Press.
- Ferman, I. (2004). The washback of an EFL national oral matriculation test to teaching and learning. In L. Cheng, Y. Watanabe & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 191-210). Mahwah, NJ: Lawrence Erlbaum.
- Hawkey, R. (2006). *Impact theory and practice: Studies of the IELTS test and ProgettoLingue 2000*. Cambridge: CUP.
- Hsu, H. (2009). The impact of implementing English proficiency tests as a graduation requirement at Taiwanese universities of technology, unpublished PhD thesis. University of York.
- Hughes, A. (1989). *Testing for Language Teachers*. Cambridge: CUP.
- Hughes, A. (1993). *Backwash and TOEFL 2000*. Unpublished manuscript, commissioned by Educational Testing Service (ETS). University of Reading.
- Markee, N. (1997). *Managing curricular innovation*. Cambridge, England: Cambridge University Press
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed., pp. 13-103). New York: ACE and Macmillan.

- Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13, 241-256.
- Morrow, K. E. (1986). The evaluation of tests of communicative performance. In M. Portal (Ed.), *Innovations in language testing*. London: NFER/Nelson [-21-].
- Pearson, I. (1988). Tests as levers for change. In D. Chamberlain & R. J. Baumgardner (Eds.), *ESP in the classroom: Practice and evaluation. ELT Documents Volume 128* (pp. 98-107). London: Modern English Publications.
- Popham, W. J. (1987). The merits of measurement-driven instruction. *Phi Delta Kappan*, 68, 679-682.
- Shepard, L.A. (1990). Inflated test score gains: Is the problem old norms or teaching the test? *Educational Measurement*, 9, 15-22.
- Shohamy, E., Donitsa-Schmidt, S., & Ferman, I. (1996). Test impact revisited: Washback effect over time. *Language Testing*, 13(3), 298-317.
- Smith, M. L. (1991). Meanings of Test Preparation. *American Educational Research Journal*, 28, 521-542.
- Tsagari, K. (2006). Investigating the washback effect of a high-stakes EFL exam in a Greek context: participants' perceptions, material design, and classroom applications. Unpublished Phd thesis, University of Lancaster
- Wall, D. (1997). Impact and washback in language testing. In C. Clapham & D. Corson (Eds.), *Encyclopaedia of language and education. Vol. 7. Language testing and assessment* (pp. 291-302). Dordrecht: Kluwer Academic.
- Wall, D. (2005). *The impact of High-stakes examinations on classroom teaching: A Case study using insights from testing and innovation theory*. Cambridge: CUP.