Washback in Language Testing: a Review of the Concept and its Implications to Innovation in Education

الأستاذ :حواجلي أحمد شوقي

قسم الآداب واللغات الأجنبية

- جامعة بسكرة-

Abstract :

Washback is a concept used to refer to a phenomenon associated with the influence language tests exert on teaching and learning. Washback can be positive, as well as it can be negative. The present article is a serious attempt in the field of testing research that seeks to shed light on the concept washback. The purpose of the article is to help practionners to get a clear and large picture on the nature, functions and mechanisms of the term under study. Besides, this presentation aims to display how this phenomenon can be a lever to provide models of innovation in education

ملخـــــص:

مصطلح «Washback » أو ظاهرة تأثير الامتحان على العملية التدريسية والتعليمية هي فعل موجود في الممارسات اليومية لمدرسي اللغات الأجنبية. هذا التأثير قد يكون له وجهان مختلفان. من جهة وجه إيجابي، ومن جهة أخرى وجه سلبي . الإشكالية المطروحة هي أنه على الرغم من وجود هذه الظاهرة إلا أن علمها مقصور على قلة منهم وكثيرهم يجهلونها. في إطار هذه الفكرة، يأتي هذا المقال لإماطة اللثام وتسليط الضوء على هذه الظاهرة من خلال توضيح طبيعتها، وظائفها، وآليات حدوثها. بالإضافة إلى هذا، يحاول هذا المقال توضيح العلاقة بين ظاهرة تأثير الامتحان أو ما يسمى «Washback»، ومجال الإبداع التربوي.

Washback is a concept now commonly used in applied linguistics. It refers to the influence of testing on teaching and learning. Language testing specialists claim that there exist two types of washback. The first type concerns the negative influence of some categories of tests, particularly, the large scale or standardized national examinations, on teaching and learning. The latter are seen to exert a harmful impact and strong pressure that often lead to limiting the scope of teaching and learning. That is, language teachers, instead of teaching a language for given purposes, turn merely to teach some tasks that are frequently included in tests and examinations their students often take for evaluation. In this way, those teachers digress from being subjects that have precise and definite tasks to fulfill to become simple teachers, whose main job is to train, and very often, in a mechanic way, their students how to respond to the typology of tasks that their tests currently comprise.

The second type of washback is the one that deals with the positive influence of language tests on teaching and learning. In this context, being aware of the power of tests, policy makers in many parts of the world continue to use these tests to manipulate the educational systems, to control curricula, and to promote new textbooks and new teaching methods. This way of conceiving things considers tests as a means that can serve its users for beneficial washback. But still, because of the intricate complexity of this phenomenon, washback in language testing remains an issue under constant debate, and a number of raised questions on this matter are yet not answered.

Therefore, in the sphere of this discussion, the present article is an additional attempt in the literature on washback that intends to review the nature, mechanisms, and pedagogical concerns and the implications of this phenomenon in educational innovation. In particular, this elucidation seeks to sensitize language teachers about the need to regard testing as an integral part of the whole teaching operation, and also to raise their attention on the crucial requirement to consider the washback effect as a phenomenon that is currently present in their daily practices, and hence strive to use beneficial washback appropriately and eliminate harmful washback.

1. **The Origin of Washback**

Although the subject of the effects of examinations has long been discussed in the literature of general education (Vernon, 1956; Kirthland, 1971; Kellaghan et al., 1982), and has been looked at from different points of view(Airasian, 1980; Popham, 1987; Madaus, 1988; Frederickson& Collins,1989; Cooley,1991; Haladyna et al., 1991), it has been common in the literature on testing that washback, as it is known for us now, has come to attract the attention of test researchers only at the beginning of the 1990's. Before that date, testing specialists and applied linguists used different terms to refer to the idea of examination influence. Some of these terms included examples such as, test impact (Bachman, Palmer, 1996; Baker, 1991), systemic validity (Frederickson& Collins, 1989), consequential validity (Messick, 1996), measurement- driven instruction (Popham, 1987), curriculum alignment (Shephard, 1993) backwash (Biggs 1996), and possibly other terms. Language testing researchers have realized that the emergence of the concept washback is the result of considerable reforms and advances that have taken part in the domain of language testing mainly during the last two decades at the end of the twentieth century.

In a comprehensive study on how the concept washback has come to emerge, Tsagari (2006) designs an artificial time framework divided into three distinct but successive phases that clearly display how the concept has evolved overtime in the scene of language testing. These phases are the "pre-1990's", the "1990's", and the "post 1990's". Tsagari identifies that the initial phase was mainly characterized as the period of time when writers recognized the existence of the examination influence phenomenon but no one accounted for it. Second, the 1990's phase was thought to be different from the previous one and was basically dominated by the publication of a seminal paper by two prominent language testing researchers, Alderson and Wall, who are greatly indebted the fact they were the first who questioned the nature of examination influence; and more importantly, they managed to re-conceptualize this phenomenon by proposing a set of hypotheses. The third phase, the "post 1990's", or as

Tsagari names it the 'reality phase', was significant since substantial models of washback have been developed in order to accurately explain and analyze the nature of this phenomenon. Such a way of looking at the subject is seen by testing experts to be a serious attempt and a step forward in the study of washback in language testing (Gosa, 2004: 29-31).

2. <u>The Definition of Washback</u>

What is notable in the testing literature on this phenomenon is that many applied linguists have indicated that the concept washback is rarely found in language dictionaries. The few available explanations can be found in examples such as the "New Webster's Comprehensive Dictionary", which defines washback as,'the unwelcome repercussions of some social actions' or another definition expressed in by the 'Collins Cobuild Dictionary of English Language', which defines "washback as 'the unpleasant after effects of an event or situation". Except these two examples, a meticulous research on this concept has shown that there does not exist an explanation or even a slight indication of washback as it is generally identified in the testing literature in the present time.

Nevertheless, unlike the rare definitions found in language dictionaries, a great deal of definitions of the concept washback is present throughout the published assessment research and literature with various meanings. For instance, Alderson and Wall (1993) define this concept as the extent to which a test influences language teachers and learners to do things. Likewise, Buck (1988) and Prodromou et al. (1995), see washback as the phenomenon that can have an influence on various aspects of teaching and learning. For Messick (1996), this term refers to the extent to which the interaction and the use of a test influence language teaching and learning to do things learners would not otherwise do that promote or inhibit language learning. Schohamy (1992) thinks that this concept is the result of the strong authority of external testing and the major impact it has on the lives of test takers. In accordance with this definition, Pearson (1988) converges with this idea and believes that washback can be a potential means for educational reforms. In another broad definition of the concept, Cheng (2005) relies on the explanation of Pearson (1988) to show the influence of the external examinations on the attitudes, behaviours,

and motivation of teachers and learners, and even on other related subjects of this research concept, and demonstrate the scope of the study, and hence make it possible to put this concept in the appropriate place for its right use.

But still, the majority of researchers in general education have reported that before yielding any explanation of the concept washback in its broadest sense, there is a crucial need to show the distinction between this concept and the other term impact in language testing. On this particular point, Tsagari (2006) argues that the common view which prevails in the domain of language testing considers washback as one dimension of impact. The latter is often used to describe effects on the wider educational context. Tsagari goes back to Wall (2005) to discuss in detail the relationship between washback and impact. For Tsagari, the former refers to the effects that tests may have on teaching and learning, whereas the latter deals with the effects that tests may have on individuals, policies, and practices within the classroom, the schools, the educational systems, or even society as a whole (p. 16). In parallel, Bachman and Palmer (1996), though they do not explicitly distinguish between the two concepts, point out that washback can be best considered within the scope of impact. Both of them think that the impact of test use operates at two levels:

- A micro-level, in terms of individuals who are affected by the particular test use, test takers and teachers, and
- A macro-level, in terms of society and educational systems(p. 17)

What sorts from the above illustration is that the concept washback can be defined according to two major perspectives. One at the narrower definition which focuses on the effects a test has on teaching and learning, and the other at a wider and more holistic view of washback that transgresses the classroom to take into account the educational system and society at large (Pan, 2009:259), which , as noted above, would be more accurately referred to as test impact. In a nutshell, Hamp-lyons (2007) summarizes the situation and the terminology well. She finds that Alderson's and Wall's limitation of the term washback to influence on teaching, teachers, and learning

seem now to be generally accepted, and the dimension of wider influences of tests is considered under the term impact, with the term used in wider educational measurement literature. In a similar view, the adoption of Bachman's and Palmer's explanation, that refers to issues of test use and social impact as 'micro' issues of impact, while washback takes place at the 'micro' level of participants, mainly teachers and learners, seem the most acceptable.

3. Types of Washback
3.1. Negative Washback

Negative washback is seen by testing researchers as the negative influence of tests on teaching and learning. Alderson and Wall (1993) refers to the negative washback as' an undesirable effect on teaching and learning of a particular test. The test may fail to reflect the learning principles and/ or the course objectives to which they are supposedly related' (p. 5). In this case, these tests will lead to the narrowing of content in the curriculum. For Vernon (1956), teachers tend to ignore subjects and activities that are not directly related to passing examination and testing accordingly after the negative curriculum in a negative way (p. 17). Again, those tests may fail to create correspondence between the learning principles and/ or the course objectives to which they should be related (Cheng, 2005:08). More dangerous, negative washback substantially reduce the time available for instruction, narrow curriculum offerings and modes of instruction, and potentially reduce the capacities for teachers to teach the content and to use methods and materials that are incompatible with useful testing formats (Smith, 1991:20). Madaus (1988) intersects with the above assumptions and points out that the negative washback definitely result in cramming, narrowing the curriculum, focus attention of those skills that are most relevant to testing, placement of constraints on teachers' and learners creativity and spontaneity, and disparage the professional judgment of educators (p. 02).

The result of negative washback is that an increasing number of coaching the classes is set up to prepare students for examination, but what students learn are test-taking skills rather than language activities (Wiseman, 1961:21). In this learning context, an atmosphere of anxiety and fear of test results becomes current among teachers and

learners. Hence, teachers will feel that success or failure of their students is reflected on them, and they speak of pressure to cover the materials for the examination. When the students know that one single measure of performance can determine their lives, they will likely to take a positive attitude toward learning.

3.2. Positive Washback

There are other testing researchers who have seen that washback in a more positive way (Andrews, Fullilove, Wong, 2002; Bailey, 1996; Davis, 1985; Hsu, 2009). Those researchers strongly assert that it is quite possible to bring about beneficial changes in teaching by changing examinations, representing the positive washback (Cheng, Watanabe& Curtis, 2004:10). This term refers to tests and examinations that influence teaching and learning positively (Alderson, Wall, 1993:15). In a broad interpretation, good tests can be utilized as beneficial teaching-learning activities so as to encourage a positive process (Pearson, 1988:07). Andrews et al., (2002) suggest deliberately introducing innovation in the language curriculum through modifications in language testing. For instance, an oral proficiency test was introduced in the expectation that it would promote the teaching of the speaking skill (Hsu, 2009:49). Davies (1985) comments on this last assumption and points out that the test no longer needs to be an obedient servant; rather, it can also be a leader.

4. The Functions of Washback

It has been stated in language testing that tests can serve a number of functions ranging from measuring students' level of competence and knowledge to imposing the effectiveness of teachers and schools. For many educators, a test is often seen as a means by which decision makers usually come to make a judgment on how instruction is carried out and learning is going on, and the extent to which the set out objectives have been attained. Traditionally, tests used to be at the end of the teaching operation to provide a diagnosis of the effects of teaching and learning. Nevertheless, with the advances and changes made in the domain of testing, and how the latter is conceived, a test

can also be developed to be used at the beginning or in the middle of the teaching/ learning process in order to influence this process and serve specific functions. This view is derived from the realization of test power and its manifestation with regard to examination decisions based on test results for individuals, educational system, and society as a whole(Hsu,2009:50).

In discussing the functions of language tests through which washback occurs in actual teaching environments, Wall(1993)refers to a number of reviews of those tests and the influence they have on the system they are introduced into. One of the most comprehensive reviews is the one that was introduced by Eckstein and Noah (1993). In its essence, Eckstein and Noah provided a historical account for the functions and influences of some examples of tests as crucial by which they take important decisions for some precise purposes. For the two researchers, the first documented use of written, public examination systems occurred under the Han Dynasty in China, about 200 B.C.

The main functions of these examinations were to select candidates for entry into the government services. In other words, the candidates were used to break the monopoly over government jobs enjoyed by the aristocratic Feudal system. Another function was to check patronage and corruption. As an example of this function was Britain where people could gain entry into higher education or the professions of strengths. An important consequence of these examinations was the establishment of numerous public schools, which aimed at preparing students for examination. The third function of examination, as presented by Eckstein and Noah, was to encourage levels of competence and knowledge amongst those who were entering government services or professions. The intention was to design examinations which reflected the demands of the target situation, and students could have to develop skills which were relevant to the work they hoped to get in the future.

 The fourth function was that of allocating sparse places in higher education. At this level, examinations were used to as a means of selecting the most able candidates for the available places. This type of examinations is quite the same to what is referred to as the placement tests in the testing literature in the present time. The fifth

function, in this illustration, was to measure and impose the effectiveness of teachers and schools. Eckstein and Noah again used Britain as an example describing how, at a certain time, the government set up a system of examinations through the allocation of considerable funds. The amount of funding that a school received depended on how its students performed. However, this system had serious unintended consequences and at least had failed to achieve the expected objectives.

The final function, in Eckstein and Noah set of examples, was limiting curriculum differentiation. In Britain, in the 19th and 20th centuries, there was a remarkable resistance to the idea of centralized education, and all schools had the freedom to decide on their own curriculum and their means of assessment. With the establishment of certificate examinations, the schools had a common target they could aim for, and all these schools turned to teach the curriculum that can help better in doing well in the examinations that are relevant to these certificates (Eckstein, Noah, 1993:5-17).

Therefore, this series of functions of tests, exposed above, are typical situations where these tests were used to exert influence- or let us say a washback effect-on the final outcomes to suit the desired intentions of those in authority to make and improve their policies. Shohamy et al. (1996) comments on this assumption and notes that 'the power and authority of tests and external examinations enable policy makers to use them as effective tools for controlling educational systems and prescribing the behaviour of those who are affected by their results, administrators, teachers, and students' (P.299). In this sense, school wide examinations are used by principals and administrators to enforce learning, which in classrooms, tests and quizzes are used by teachers to impose discipline and motivate learning. Given this status of tests and public examinations, a systematic study of the functions of tests in learning and teaching is essential.

5. The Mechanisms of Washback

In exploring the complex mechanisms through and by which washback occurs in actual teaching and learning environments, Bailey

(1996) cited Hughes (1993) trichotomy to show how this phenomenon works in different contexts. Bailey points out that this trichotomy allows educators in education in general and testing specialists in particular to develop a basic model of washback that explains how the various components that make-up this framework interact to help in understanding the nature of this subject of interest. In describing this particular model, Hughes states that the trichotomy is formed of three parts, first, the participants who are mainly the people such as students, classroom teachers, administrators, materials developers and publishers whose perceptions and attitudes toward this work may be a test. Hughes second component in this framework is termed process. The latter covers any actions taken by the participants which may contribute to the process of learning as the development of teaching materials. Third, in Hughes' framework, a product refers to what is learnt as facts, skills, and other aspects and the quality of learning (p. 02).

Hughes goes on to make this model clear enough for all people who are concerned with the mechanisms of washback and further points out the following elucidation on this particular issue. He states that

> The trichotomy into participants, process, and product allows constructing a basic model of washback. The nature of a test may first affect the perception and attitudes of the participants towards their teaching and learning tasks. The perception and attitudes in turn may affect what the participants do in carrying out their model(process), including practicing the kind of items that are to be bound in the text, which will affect the learning outcomes, the product of that work(p.02).

Indeed, this elucidation in which Hughes makes a clear distinction between the three components of this model, and where he stresses on the participants' perceptions and attitudes, and how these factors affect what they do is a comprehensive presentation where a test not only affects the three components, but also provides feedback on how washback functions.

Contrary to Hughes who stresses more the three components that make up this model/framework, Alderson and Wall(1993), in another study, focus on what they referred to as the 'micro-aspects' of teaching and learning that might be influenced by examinations(Cheng, Watanabe& Curtis, 2004). They argue that there is little evidence provided by empirical research to sustain the idea that tests impact on teaching. They advocate that 'the concept is not well defined, we believe it is important to be precise about what washback might be before we can investigate its nature and whether it is a natural or inevitable consequence of testing' (p.117). As a matter of fact, they suggest 15 hypotheses that can aid researchers to illustrate areas of teaching and learning that are usually affected by washback, and can stand as a basis for further researches (Cheng, 2005). The 15 hypotheses are stated as follows

1) A test will influence teaching;

2) A test will influence learning;

3) A test will influence what teachers teach; and

4) a test will influence how teachers teach.

5) A test will influence what learners learn; and

6) a test will influence how learners learn;

7) A test will influence the rate and sequence of teaching; and

8) A test will influence the degree and the depth of learning;

9) A test will influence the degree and the depth of teaching and

10) a test will influence the degree of learning.

11) A test will influence attitudes towards the content and method of teaching and learning.

12) Tests that have important consequences will have washback; and conversely

13) tests that do not have important consequences will have no washback.

14) Tests will have washback on learners and teachers.

15) Tests will have washback effects for some learners and some teachers, but not for others.

In sum, the above discussion of the common studies on the mechanisms of washback has indicated that there exist a strong tie

between the subject of test design and its impact and power on teaching and learning either positively or negatively. However, it has been stated that even if these studies have contributed in advancing research into the domain of washback in language testing, but still they remain inefficient to draw a larger and clearer picture of this issue since a number of raised questions on the mechanisms of washback in language testing remain unanswered.

06. Washback- a Phenomenon Leading to Educational Innovation

It is assumed that to understand the nature of washback, it is also crucial to take account of findings in the research literature in the area of innovation in language and change in educational settings. The reason of this view is that many applied linguists consider that there are many ideas in educational innovation which can stand as a solid ground for language testing specialists to judge whether the tests they are designing are likely to have the impact they intend them to possess (Wall, 2005). On this particular point, Hsu (2009) asserts that there has been a well-established tradition, which led to the realization of a number of networks that served to yield the most elegant compilations of assumptions about the different phases in the innovation process at the factors at work in every phases (Fullan, 2007; Rogers, 2003), and an increasing body of literature focusing on the English language teaching context (Henrichsen, 1989; Kennedy, 1990; Li, 2001; Markee, 1993; Stoller, 1994; White, 1993). The particularity of these studies is that they managed to clarify the complexity of the innovation process, and the factors which inhibit or facilitate successful diffusion and implementation.

Following Wall(2005), Rogers(2003) defines innovation as an' idea, practice, or object that is perceived as new by an individual or other unit of adoption'(p.11). In Hsu (2009), innovation can be usefully defined as a planned and deliberate effort, seen as a new act by an individual or individuals to bring about improvement in relation to desired objectives. Hsu makes this last assumption more explicit. He advocates that educational innovation is the result of a number of problems that a given educational system can present as a failure of students' achievement, a poor performance by students in specific

areas, or lack of transparent accountability reporting. What is significant about these problems is that the latter also transgress to touch some aspects of educational system that concern systematic attempts by some authorities to change educational policies with the intention to achieve better outcomes.

On the ground of this elucidation, a number of models have been provided to make the subject of innovation in language education more practical and possibly easy to incorporate in acts that intend to yield desirable changes or as it is proposed in the domain of testing, a factor leading for washback. For instance, Rogers (2003) discusses the idea of diffusion and posits that this process involves the communication of a new idea in which those who are involved, or as it is often referred to as participants, create and solve this new idea with the expectation to attain a certain natural understanding. For Roger's, diffusion is a sort of a ' social change' by which this phenomenon happens in a social function. In other words, diffusion is the process by which washback is communicated through certain ways among the members that make-up a social system. Moreover, in a comprehensive survey of innovation in education, Fullan (2007) regards the issue of innovation as a process rather than as an event. He shifts the setting of educational changes to a broader stage, introducing these phases in the innovation process: initiation, implementation, and continuation.

- The initiation stage: it is the process that occurs between the first appearance of the idea for change and the time when it is adopted.
- The implementation stage: it is the process of putting into practice an idea, programme, or a set of activities and structure new to the people attempting or expected to change.
- The continuation stage: it refers to whether an innovation becomes part of the educational system, or whether it fails and/or is rejected.

In addition to these two models, Markee's (1997) framework in the area of educational innovation is regarded by a great deal of researchers to be among many other models that successfully realized to summarize and display the relevance of the idea from innovation

theory. In its essence, Markees' model provides a set of principles for language teaching professionals to understand the facts that affect the design, implementation, and maintenance of innovation. In explicit terms, this model is organized around answers to the questions that originally were posed by Cooper (1982), and that include questions as 'who adopts, what, when, and why?'

- Under 'who', Markee introduces a description of the participants in the innovation process.
- Under 'what', Markee defines innovation as an idea perceived as new by individuals which is intended to bring about improvement in relation to desired objectives.
- Under 'where', Markee stresses the importance of understanding the context where the issue of innovation takes place.
- Under 'when', Markee discusses the rate of diffusion
- Under 'why', Markee discusses the characteristics of adopters and features which can facilitate or hinder innovation.
- Under 'how', Markee describes aspects affecting change.

If it is applied to language testing, Markee's model would help testing specialists to realize whether washback has occurred or not. And more importantly, the nature of washback can be displayed.

The final framework in this series of innovation models is the one of Henrichsen. In its basics, Henrichsen's model illustrates how the many factors in the diffusion work and how they interact with one another. He claims that the common criteria of this model are coherence, abstractness, completeness, and relevance to contact change, and cross-cultural applicability. On the basis of this view, Henrichsen sees that the diffusion process has three components: 'Antecedents', 'Process', and 'Consequences'.

- The 'Antecedents' component: it is the set of conditions of the educational context or environment before an innovation is introduced.
- The 'Process' components: it is concerned with the factors that stand as facilitators and/ or hinders to change.
- The 'Consequence' component: it is the description of how a decision to be adopted or rejected as an innovation.

In light of the above description of Henrichsen's model, the latter can be very helpful in explaining the process by which innovations are either accepted or rejected by their intended receivers. And in the mean time, it is appropriate to judge whether washback occurs, or not.

In a nutshell, the purpose of eliciting this literature is to provide enough information on the intricate relationship between innovation in education and washback in language testing. Thus, it is evident that different theories of innovation and change have yielded insights on how researchers should provide to implement subjects that are new for the people concerned by this change. Besides, an understanding of the basics of these described models gives a better interpretation of the nature of washback, and more importantly, how this latter works when time comes to implement an innovative act.

Conclusion

To conclude, this article has reviewed a number of issues related to the subject of washback. In precise terms, an attempt has been made to shed some light on the origins, definition, and mechanisms of this phenomenon. What is notable is that this concept is central to intricate relationships between testing, teaching and learning. The present elucidation has considerably helped us to display the power and authority of tests on teaching and learning, indicate how tests become effective ways for influencing educational systems, and prescribe the behaviour of those who are affected by their results. 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. In parallel, those studies have shown that a few of them are tackled by practionners in the local scene, and yet this subject is not well considered and exploited to the extent that is supposed to go in pertinent accordance with research in other areas in the field of language teaching methodology in Algeria. Thus, this paper can stand as a step towards further researches in language testing.

References

- Airasian, P.W (1980). Issues and misuses of standardized
        Measurement in Education, 10, 1-8.
- Alderson, J. C., & Wall. D. (1993). Does wash exist? Applied
        Linguistics, 14, 15-129.
- Andrews, S., Fullilove, J & Wong. Y. (2002). Targeting washback: a
        case study. System, 30 (2), 2007- 223.
- Bachman, L. F., & Palmer, A. S. (1996). Language testing in practice.
        Oxford University Press.
- Baker, E. L. (1991). Issues in policy, assessment in equity. Paper
        presented at National Symposium on limited English
        proficient students' issues: focus on evaluation and
        measurement, Washington, D. C.
- Bailey, K. M. (1996). Working for washback: a review of the
        washback concept in language testing. Language
        Testing, 13, 257-279.
- Biggs, J. B. (1996). Testing to educate or to select? Education in Hong
        Kong at the cross- roads. Hong Kong: Hong Kong educational
        publishing.
- Buck, G. (1988). Testing listening comprehension in Japenese
        University entrance examinations. JALT Journal, 10, 12-42.
- Cheng, L. (2005). Changing language teaching through language
        testing: Cambridge: Cambridge University Press.
- Cheng, L., Watanabe, Y. & Curtis, A. (2004). Washback in language
        testing: research contexts and methods. Mahwah, NJ: Lawrence
        Erlbaum Associates.
- Cooley, W. W. (1991). State- wide student assessment: educational
        measurement: issues and practices, 10, 3- 36.

- Cooper, R.L. (1982). Language spread: studies in diffusion and social
        change. Bloomington: IndianaUniversity Press, and D. C.,

Center for Applied Linguistics.

- Davies, A. (1985). Follow my leader: is that what language tests do? In Y. P. Lee, A. C. Y. Y. For. Lord& G.Low (Eds), New directions in language testing, 3(12). Oxford: Pergamon Press.
- Enkstein, M. A. & Noah, H. J. (1993). Secondary school examination: international perspectives on policies and practices. New Haven: Yale University Press.
- Frederickson, J. K. & Collins, A. (1989). A system approach to educational testing. Educational Research, 18(9), 27- 32.
- Fullan, M. (2007). The New meaning of educational change (4th Ed). New York: Teachers College.
- Gosa, C. M. C. (2004). Investigating washback: a case study using diaries. Unpublished PhD thesis, Departments of linguistics and English language, University of Lancaster.
- Haladyna, T. M., et al. (1991). Raising standardized achievement tests scores and the origin of test score pollution. Educational Research, 20(5), 2-7.
- Hamp-Lyons, L. (2007). The Impact of testing practices on teaching: ideologies and alternatives. International Handbook of English language teaching, 1, 487-504, New York: Springer .
- Henrichsen, L. E. (1989). Diffusion of innovation in English language teaching: the ELEC effort in Japan, 1956- 1968. Connecticut: Green Wood Press.
- Hsu, H. F. (2009). The impact of implementing English proficiency tests as a graduation requirement at Taiwansese universities of technology. Unpublished PhD thesis, University of York.
- Hughes, A. (1993). Backwash and TOEFL 2000. Unpublished Manuscript, University of Reading.
- Kellaghan, T., Madaus, G. F. & Airasian, P. W. (1982). The Effects of standardized testing. London. Kluwer Nijboff Publishing.
- Kennedy, C. (1990). Evaluation of the management change in ELT projects. Applied Linguistics, 9 (4), 329- 342.
- Kirthland, M. C. (1971). The Effect of tests on students and schools. Review of Educational Research. 41 (4), 303- 350.
- Li, D. (2001). Teacher's perceived difficulties in introducing the communicative approach in South Korea, In D. Hall, & A.

Hewings (Eds). Innovation in English language teaching, 149-169.   London: Routledge

- Madaus, G. F. (1988). The Influence of testing on the curriculum. In Tanner, L. N. (Ed.). Critical issues in curriculum: Eight Seventh Yearbook of the National Society.

- Markee, N. (1993). The Diffusion of innovation in language teaching. Annual Review of Applied Linguistics, 13, 229- 243.

- Markee, N. (1997). Managing curricular innovation. Cambridge: Cambridge University Press.

- Messick, S. (1996). Validity and washback in language teaching, Language Testing, 13 (3), 24- 256.

- Pan, U. C. (2009). A Review of washback and its pedagogical I mplications. In VNU journal of Sciences, Foreign Languages (25), 257- 263. Australia: University of Melboune.

- Pearson, I. (1988). Tests as levers for change. In D. chamberlain & R. J. Baungardner (Eds.), ESP in the classroom: practice and evaluation 98, 107. Great Britain: Modern English Publication.

- Popham, W. J. (1987). The merits of measurement- driven instruction. Phi Delta Kappa, 68, 679- 682.

- Prodroumou, E., Donista, Schmidt, S., & Ferman, I. (1996). Test impact revisited: washback effect over time. Language Testing, 13(3), 298- 317.

- Rogers, E, M. (2003). The diffusion of innovation in language teaching. Annual Review of Applied linguistics, 13, 229- 243.

- Shephard, L. A. (1993). The place of testing reform in educational reform: A reply to cizek. Educational Research ,22(4),10-14

- Shohamy, E. (1992). Beyond the proficiency testing: a diagnostic feedback testing model for assessing foreign language learning. The Modern Journal, 76(4), 513- 521.

- Shohamy, E., Donista- Schmidt, S., & Ferman, I. (1996). Test impact revisited: washback effect over   time. Language Testing, 13(3), 298-317.

- Smith, M. L. (1991). Put to the test: the effects of external testing on, teachers. Educational Research. 20(5), 8- 11.

- Stoller, F. L. (1994). The diffusion of innovation in intensive ESL

progress. Applied Linguistics, 15(3), 300-327.
University of Lancaster

• 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. (2005). The impact of high-stakes examination on classroom
teaching: a case study using insights from testing and
innovation theory. Cambridge: Cambridge University Press.

• Wall, D., & Alderson. J. K. C. (1993). Examining washback: the
Srilankan impact study. Language Testing, 10, 41- 70

• White, R, U. (1993). Innovation in curriculum planning and program
development. Annual Review of Applied Linguistics, 13, 244-
259.

• Wiseman, S. (1961). Examination and English education. England:
Manchester University Press.

• Vernon, P. E. (1956). The Measurement of abilities (2nd Ed.). London:
University of London Press.