

Gateways to Positive Washback of Language Testing: A Review of the Factors that Contribute to Washback Effects

Asiqur Rahman¹

Abstract

The ways language testing influence the process and scope of language teaching and learning are generally known as washback or backwash effects. Washback may be positive or negative depending on multifarious factors. However, this study aims to identify the most effective ways to maximize the positive washback and minimize the negative washback of language testing. Drawing on a body of washback studies, this qualitative research concludes that the most vital factors that lead to positive or negative washback are related to the test types, process of test development, the ways of scoring, and socio-economic influences

Keywords

Washback, test types, test development, representative sampling, task-based, criterion referenced

Introduction

There is a consensus that assessments directly influence educational processes in several ways. Although good education focuses more on learning and less on testing, hardly anybody can deny that assessment is an inseparable part of language teaching/learning. Conventionally, students, apart from the learning of lessons, take tests, which can be either formative or summative. So, teachers usually end up teaching to the tests. Thus, testing is a significant part in the process of language learning and teaching. And naturally these tests have a significant impact on teaching and learning which is known as washback or backwash. Numerous studies have found that tests influence teaching content, course design and classroom practices (Cheng, 2005; Cheng, 2007; Green, 2007; Pan, 2009; Shih, 2010; Wall, 2005). Heaton (1988, p. 5) states, “Both testing and teaching are so closely interrelated that it is virtually impossible to work in either field without being constantly concerned with each other.”

¹Lecturer, Department of English, School of Liberal Arts and Social Sciences (SLASS), Chittagong Independent University, Chattogram, Bangladesh

Corresponding author:

Asiqur Rahman, Lecturer, Department of English, School of Liberal Arts and Social Sciences (SLASS), Chittagong Independent University, Chattogram, Bangladesh

Email: asiqur@ciu.edu.bd

Therefore, there is, undoubtedly, a close tie between language testing and language teaching pedagogy. Washback, the influence of language tests on language teaching and learning, can be either positive or negative depending on certain factors.

It is generally believed that a good test can widen the scope of teaching and learning which is called positive backwash, while a bad test can limit the scope of teaching and learning that is called negative washback. Thereby, it is important for the language teachers to know how to develop good tests for the learners. However, two basic questions – what is a good/bad test in terms of washback? Is it possible to set some principles which will ensure beneficial washback and help avoid harmful washback? – remain unanswered. This study explores the most vital factors (around these questions) that contribute to beneficial or harmful washback.

Literature Review

The term backwash or washback has been defined in several ways. Generally, the effects of tests on teaching and learning are obvious in language teaching and learning. According to Arthur Hughes (2003), the effect of testing on teaching and learning is known as backwash. He adds that preparation for a test can come to dominate all teaching and learning activities. According to Davies (1968), a good test is like an obedient servant of teaching. Although Davies considers the relationship between teaching and testing as that between servant and master, Hughes (2003), on the other hand, prefers to call it a partnership. According to Hughes (2003), when the teaching program is potentially good and appropriate, and the testing is the opposite, it is very likely that learning and teaching suffer from harmful backwash. Shohamy (1992) mentions about the importance of proper use of external language tests in influencing foreign language learning in the school context. Biggs (1995) asserts that the curriculum, pedagogy and the learners' learning scopes are driven by testing. Again, Messick (1996) defines washback as what language teachers and learners are to do influenced by tests. He adds that the evidence of teaching and learning effects should be understood as washback as long as that evidence is related to the introduction and use of the test. As wider and more complex aspects of washback, Ali and Hamid (2020) point out the significance of socio-political and -economic factors that contribute to negative washback.

Washback effects can be positive in certain situations. Hughes (2003) emphasizes that there will be positive washback effects when test samples are wide and unpredictable; when the test is a direct one; when test scoring is criterion referenced which analyzes and identifies learners' linguistic abilities and limitations; and when the test is based on course objectives. Almond, Mislavy and Steinberg (2002) suggests Task-based language assessment (TBLA) which is likely to bring a positive backwash. Bailey (1996) implies that beneficial washback can be promoted by incorporating language learning goals, authenticity, learner autonomy and self-assessment, and detailed score

reporting into our tests. Bailey (1996) also lists other criteria likely to promote beneficial washback, such as:

1. Test-takers, teachers, administrators, and curriculum designers should understand the purpose of the test;
2. Results must be believable to test takers and score users;
3. Test takers must find the results credible and fair; and
4. Test should measure what the programs intend to teach.

Wall (1996), in reviewing the literature, lists the desirable characteristics in language testing as being the following (p. 33a):

1. Direct testing,
2. Criterion-referencing,
3. Authentic texts,
4. Tasks.

Weir (1993) suggests that communicative language testing could have a strong washback effect on communicative language teaching and, in fact, that such a washback effect would be directly linked to the construct validity of the tests.

Nevertheless, washback may be harmful under certain circumstances. Lynda Taylor (2005) contends that negative wash back is likely to occur by the constrictions of the teaching and learning context when a test's content or format is founded on a narrow language ability. For example, when writing skill is assessed only with Multiple Choice items, which would lead the students and teachers to devise tricks to handle MCQs only, rather than practicing actual writing skill itself. As according to Hughes (2003), an indirect test like MCQs for testing writing skill is likely to bring harmful backwash. Wall and Alderson (1996), as cited by Ali and Hamid (2020), note that negative washback may occur in two ways: (a) teachers may focus only on the content that is relevant for examination, avoiding other content in the textbook; and (b) the content of the test may not represent the textbook, making the test unrelated to the curriculum. Wall and Alderson (1996) observe that negative washback can be located in three areas: teaching content, method and assessment. Ali and Hamid (2020, p. 142), in exploring the external factors contributing to washback, says,

Social forces, including the expectations of high grades, the negotiation between publishers and the teachers, the market demand of exam-preparation guide-books, inadequate supply of skilled teachers, and the unequal distribution of economic resources across schools all contribute to what is tested and taught in the school domain.

Ali and Hamid (2020) also point out that harmful washback can lie within the system and that this testing-teaching relationship may have an ideological character.

The above literature illustrates multifarious factors which are responsible for positive or negative washback, while this paper calls attention to the most agreed-upon factors that contribute to positive or negative washback of language testing.

Research Questions

In the light of the literature discussed above, it is clear that backwash effects might be either harmful or beneficial depending on diverse factors. This research aims to investigate and find out the most effective ways to achieve beneficial wash back and to avoid harmful backwash, and thereby ask the following questions:

1. Which are the key factors that can yield to positive washback effects of language testing?
2. What are the main pitfalls in language assessments that result in negative washback?

Method

This secondary research follows a qualitative approach. It closely reviews and analyzes the existing theories on washback and summarizes factors those contribute to both the positive and negative washback.

Findings

By reviewing the existing literature regarding washback, it is found that washback can be beneficial or harmful depending on four major factors: i) types of tests ii) how a test is developed iii) the ways of scoring /evaluation and iv) the socio-economic influences.

The table below summarizes the factors which promote positive washback as well as the ways to avoid negative washback as suggested in the literature discussed above.

Table 1. Factors that Promote Positive Washback and Ways to Avoid Negative Washback

Types of washback	Types of language tests	Process of test development	Scoring /evaluations	Socio economic factors
Beneficial /positive	Direct & Task Based testing	Wide, unpredictable and representative, sampling, course objective-based sampling	Criterion referenced, and objective	Tendency to learn and improve
Harmful /negative wash back	Indirect Testing	Narrow, unrepresentative, predictable, and class lesson-based sampling	Norm referenced, and subjective	The pressure of passing with better grades

Source: Literature Survey

Discussion

The second row of the table represents positive washback factors while the third row shows the factors that contribute to harmful washback. However, the four main factors influencing backwash, as shown in the table above, are discussed below:

1. Types of Tests

Washback depends on the types of language tests to a great extent. Most importantly, a direct and task-based test is likely to bring beneficial washback, while an indirect test results in harmful washback. An indirect writing test, for instance, does not require students to produce any piece of writing; rather, they are asked to choose the correct grammar or right vocabulary in an MCQ test. This kind of indirect testing is based on the principle that if a learner is able to answer MCQs or gap filling questions on items such as vocabulary, grammar, and punctuation, they will also be able to write a composition, which anyone can easily doubt. It is very likely that this kind of test will lead to negative washback since as an influence of this type of test, learners and teachers automatically get busy devising the tactics of finding out the correct answers, rather than genuinely developing their actual writing skills. Also, some obvious writing skills such as coherence and cohesion, paragraphing, and idea development are likely to be ignored by both learners and teachers as an impact of this type of indirect writing tests. On the contrary, learners and teachers are bound to learn and practice genuine writing skills when they are to prepare to appear in a direct writing test which involves writing a composition or paragraph. If we consider a pronunciation test, a direct test requires test takers to speak during the test, while an indirect test requires them to write an essay on the features of pronunciation or to transcribe certain words, or phrases by using IPA symbols. Therefore, while preparing for a direct pronunciation test, teachers and students necessarily emphasize on speaking practice using all the features of pronunciation ranging from pronunciation of individual phonemes to intonation, meaningful, chunking, stress timing and rhythm. Conversely, while preparing for an indirect pronunciation test, students and teachers usually remain busy memorizing the theories of productions as well as developing the ability to do transcriptions, which ultimately limits the scope of learning and thereby leads to negative washback. Thus, it can be said that a direct test, by principle, is advisable for achieving beneficial backwash.

2. The Process of Test Development

Washback also depends on the process of test development. Most significantly, a wide selection of test items taking representative samples is likely to have a good effect on learning and teaching because it automatically widens the scope of learning and teaching. This is a significant part of the construct validity of the test. For example, if in a speaking test, the examiner asks questions ranging from present, past and future tense, the learners are to practice speaking in past, present and future tense as part of their preparation for the test. Therefore, every question asked should be a representative sample relating to different

language properties in order to picture the underlying construct of the candidates, which significantly widens the scope of both learning and teaching. If a writing test includes different genres of writing, learners and teachers are to learn and practice a variety of writing genres as part of their test preparation. Nevertheless, if the writing test always includes one or two types of writing – for example, an argumentative essay --

learners and teachers usually end up working on only one type of writing, which again limits the scope of learning. Hence, the wider is the sampling, the better will be the washback. Besides, through a wide sampling, a test has to be made as unpredictable as possible for achieving beneficial washback. This is because, if the test samples are predictable to the students, they generally try to memorize predefined answers instead of trying to develop their skills. For example, if they can predict what essays they will have to write, that stops them from learning and practicing writing skills. They, instead, find it better to start memorizing ready-made essays, which is a harmful backwash. Lastly, for achieving a positive washback, a test has to be based on course objectives rather than the teaching contents. Wide sampling and unpredictability towards good washback cannot be implemented unless the test is based on course objectives rather than the lecture contents. Objectives of a language course or syllabus are usually wide and ambitious. But due to time limitation and other factors, lecture contents may not be always as wide as the course objectives. Now, if the test is based on the lecture content, it, to a great extent, shrinks the scope of learning and thereby results in negative washback. This is for the fact that, if learners already know that their test will be based on what their teachers teach in the classes, they will not learn widely. Conversely, basing achievement tests on course objectives will widen the scope of learning since usually the objectives of a language course are usually wide and ambitious. Course objectives usually state what is expected from the learners at the end of the course. For example, if the objective of a reading course is that at end of the course students are expected to be able to read 500 WPM with 70% comprehension rate, learners and teachers will try to really develop their reading speed and comprehension rate rather than trying to master the example reading passages taught in the classes and given in the textbook and adopt some cheap exam tricks and techniques of handling different types of reading exercise question like, MCQs, True / False, and Matching. Therefore, tests based on course objectives should bring better backwash on both teaching and learning.

3. Scoring/ Evaluations

Backwash also depends, to an extent, on the process of test scoring or evaluation process. Firstly, a criterion referenced test, rather than a norm-referenced test, is likely to lead towards better washback. While a norm-referenced evaluation gives students only their grades/levels/standings, a criterion referenced evaluation analyzes learners' linguistic abilities and limitations

tions. So, a criterion referenced evaluation helps teachers better make their teaching materials and methods determining the actual needs of the students. Students can also find out the specific areas to work on and improve. Thus, a criterion referenced backwash has a beneficial washback effect on teaching and learning. This criterion referenced evaluation is more possible in continuous assessment or in formative achievement tests rather than in the final or summative achievement tests. This is because formative achievement tests take place throughout the semester where corrective feedback can be given and worked on whereas summative or final achievement test take place at the end of the semester where feedback cannot be worked on. Hence, continuous of formative assessments should carry more weight in order to gain positive washback. Secondly, objectivity and subjectivity of scoring matter in gaining positive washback. Evaluation of receptive skills (listening & reading) is usually objective whereas that of productive skills (speaking & writing) is often subjective. To make the evaluation of productive skills as objective as possible, it is important to set detail marking criteria or rubrics for the examiners. If it can be done, then examiners can follow marking criteria while scoring, instead of their subjective judgment. And when the detail marking criterion or rubric is made open for students and teachers, it brings beneficial washback in the sense that once teachers/learners get access to the marking criteria or rubric, they will try to develop the language skills in the right direction provided the marking criteria or rubrics are set properly.

4. Socioeconomic Factor

Another factor that might influence washback is socioeconomic mind set of learners and parents. If, for socioeconomic reasons, learners remain under pressure to pass with higher grades, rather than widening the horizon of learning, it becomes very challenging for teachers to implement the actual course objective to the fullest extent. Teachers feel a kind of pressure to compromise and limit the course content and make it easy for the learners to pass and get grades. And it can also be considered as negative wash back in the wider and more complex sense. Motivating the students towards the development of their actual language skills and its benefits could be a good solution here.

Conclusion

By reviewing the relevant literature, this study concludes that positive/negative washback mainly depends on four main factors: types of language tests, the process of test development, scoring/evaluations and the socio-economic factors. Therefore, for maximizing the positive wash back and minimizing negative washback, teachers, educators, planners and other stakeholders must have assessment literacy regarding the four factors mentioned above. The factors identified above are general aspects of washback. However, factors contributing to washback may vary depending on a number of factors such as the types of tests, levels of test takers, types of skills tested and the importance of the test. More specific case studies can be conducted to explore these varying complexities of washback.

References

- Ali, M. M., & Hamid, M. O. (2020). Teaching English to the test: Why does negative washback exist within secondary education in Bangladesh? *Language Assessment Quarterly*, 17(2), 129-146.
- Almond, R., Steinberg, L., & Mislevy, R. (2002). Enhancing the design and delivery of assessment systems: A four-process architecture. *The Journal of Technology, Learning and Assessment*, 1(5).
- Bailey, K. M. (1996). Working for washback: A review of the washback concept in language testing. *Language Testing*, 13(3), 257-279.
- Biggs, J. (1995). Assessing for learning: Some dimensions underlying new approaches to educational assessment. *Alberta Journal of Educational Research*, 41(1), 1-17.
- Cheng, L. (2007, November). What does washback look like? In *The Sixteenth International Symposium on English Teaching*, Taipei.
- Cheng, L. (2005). *Changing language teaching through language testing: A washback study*. Cambridge: Cambridge University Press.
- Davies, A. (1968). *Language testing symposium: A psycholinguistic approach. Language and language learning [series], no. 21*. New York: Oxford University Press.
- Green, A. (2007). Washback to learning outcomes: A comparative study of IELTS preparation and university pre-session language courses. *Assessment in Education*, 14(1), 75-97.
- Hughes, A. (2020). *Testing for language teachers*. Cambridge: Cambridge University Press.
- Heaton, J. B. (1988). *Writing English language tests*. New York: Longman Inc.
- Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13(3), 241-256.
- Shih, C. M. (2010). The washback of the General English Proficiency Test on university policies: A Taiwan case study. *Language Assessment Quarterly*, 7(3), 234-254.
- Shohamy, E. (1992). Beyond proficiency testing: A diagnostic feedback testing model for assessing foreign language learning. *The Modern Language Journal*, 76(4), 513-521.
- Taylor, L. (2005). Washback and impact. *ELT Journal*, 59(2), 154-155.
- Wall, D. (2005). *The impact of high-stakes examinations on classroom teaching: A case study using insights from testing and innovation theory* (Vol. 22). Cambridge: Cambridge University Press.
- Wall, D. (1996). Introducing new tests into traditional systems: Insights from general education and from innovation theory. *Language Testing*, 13(3), 334-354.
- Wall, D., & Alderson, J. C. (1993). Examining washback: The Sri Lankan impact study. *Language Testing*, 10(1), 41-69.
- Weir, C. J., & Weir, C. J. (1993). *Understanding and developing language tests*. New York: Prentice Hall.
- Yi-Ching, P. (2009). The impact of test design on teaching. *The International Journal of Educational and Psychological Assessment*, 3, 94-103.