

# The Impact of Language Testing Washback in Promoting Teaching and Learning Processes: A Theoretical Review

Faten Alqahtani<sup>1</sup>

<sup>1</sup> English Department, Faculty of Languages and Translation, King Khalid University, Abha, Saudi Arabia

Correspondence: Faten Alqahtani, English Department, Faculty of Languages and Translation, King Khalid University, Abha, Saudi Arabia.

Received: April 15, 2021

Accepted: May 31, 2021

Online Published: June 2, 2021

doi: 10.5539/elt.v14n7p21

URL: <https://doi.org/10.5539/elt.v14n7p21>

## Abstract

Existing literature indicates that assessment is a critical aspect of teaching and learning language; the outcomes of testing are vital. The history of assessment can be traced back to when exams primarily served two significant purposes in China: choosing candidates for admission into government offices and preventing corruption. Washback as a concept can be traced back to the 1990s. It was advanced by Alderson and Wall in 1993 as a force that obliges test-takers and tutors to engage in particular tasks or activities due to exams. In this regard, washback is an impact that a test has on the teaching and learning process. High-stakes exams like the LOBELA demonstrate the significance of washback in the Saudi English-as-a-foreign-language context. This paper explores the mechanisms through which washback occurs in teaching and learning processes, ways to determine its validity, and different types of washback. It further highlights the impact of washback in promoting teaching and learning processes, as well as the role it plays in policy development in the educational system.

**Keywords:** washback, LOBELA, language testing, assessment, teaching and learning, Saudi Arabia

## 1. Introduction

Throughout history, the administration of tests to the public has played a vital role in social development. Tests were originally used in China to select officials to occupy public offices and control corruption (Cheng, 2000). Currently, examinations are used to measure students' abilities and understanding of course content. Most importantly, examinations play a critical role in educational improvement. In this regard, tests enhance accountability and reinforce the teaching and learning processes. Furthermore, the advent of globalization has facilitated the internalization of higher education. The emergence of international students worldwide has created the need to use language instruction tests, such as English tests, as a standard for language testing. However, there are concerns among scholars regarding the impact of such tests on testers and test-takers; it has been proposed that skills in a given language may suffer as a result of these tests (Shirzadi & Amerian, 2020; Rea-Dickins & Scott, 2007). As a result, different studies have also examined how tests can influence classroom management, teaching materials, and policy development (Al Hinai & Al Jardani, 2020; Spratt, 2005; Syafrizal & Pahamzah, 2020). In applied linguistics, the effects or consequences of testing are regarded as washback. This paper examines various studies to elucidate the definitions, origin, models, and types of washback, as well as show its impact on teaching and learning

## 2. Concept of Washback

Cheng (2000) describes washback as the impact a test has on the learning process. It is the consequences or effects of testing (Rea-Dickins & Scott, 2007). This means that a test impacts both teachers and students. Impact here refers to influence (Al Hinai & Al Jardani, 2020). Different scholars have suggested other definitions in straightforward and complex terms in attempts to explain washback. Spratt (2005) asserts that washback determines what test-takers are taught, the way they are taught, what they learn, and the way they learn. In this context, it is evident that washback involves a test's side effects on the teaching and learning processes (Syafrizal & Pahamzah, 2020). Alternatively, Beikmahdavi (2016) conceptualizes washback as an intricate phenomenon that surpasses the influence a test might have on the processes of teaching and learning. Other research has gone further and described washback as a highly complex term that has found no consensus. That said, to effectively describe washback, there is a need to study this multidimensional term in relation to the different effects it has on various stakeholders (Shirzadi & Amerian, 2020; Beikmahdavi, 2016).

Higher learning in Saudi Arabia is based on outcome-based-education, which is used as a benchmark against which students' success may be evaluated. The English language is incorporated in the preparatory year program (PYP) in Saudi higher learning institutions. This program is critical as it determines what university programs students enroll in, directly affecting their future careers. In this case, teaching staff is mandated to effectively prepare students to enhance their chances of excelling in the PYP tests, which are high-stakes exams. Due to the mounting pressure to produce good results and the fact that PYP exams act as a means of assisting students in fulfilling their goals, some lecturers end up teaching the exams' content, which results in harmful washback. In this regard, it is normal for lecturers to exploit all necessary channels to guarantee that students excel in the exams. Hazaea and Tayeb (2018) established that exams might influence teaching mechanisms, attitudes, motivation, and content assessments. They used the Learning Outcome Based English Language Assessment (LOBELA) to evaluate the impacts of washback on these factors. The LOBELA is a high-stakes exam in Saudi Arabia, exerting significant pressure on lecturers. Hazaea and Tayeb (2018) affirm previous studies' assertions that exams significantly affect what is taught to learners and how it is taught (Al Hinai & Al Jardani, 2020; Spratt, 2005; Syafrizal & Pahamzah, 2020). Saudi Arabia adopts a test-driven educational approach. Accordingly, Hazaea and Tayeb (2018) discovered that the LOBELA critically affects the education settings' teaching mechanisms, teaching staff attitudes and motivation, and the content of assessments.

Regarding teaching mechanisms, the authors established that the LOBELA has obliged teaching staff to adopt test-oriented teaching methodologies. This includes incorporating the exam's format in classroom work, putting extra effort into teaching vocabulary and grammar, as well as integrating new teaching approaches to improve students' understanding of the learning content. In some cases, teaching staff have embraced quizzes to enable learners to adequately familiarize themselves with exams. All these efforts are geared toward ensuring that students have higher chances of excelling in the formal exams. However, the exam's effect on teaching mechanisms has engendered some adverse effects (negative washback) on the curriculum. These include limiting the curriculum, restricting teaching approaches, and increasing pressure on teaching staff (Hazaea & Tayeb, 2018). Additionally, the LOBELA was found to affect teaching staff's attitudes and motivations. In this regard, since the exam is so important, it considerably affects teachers' perceptions of how they teach the exam's content. Lecturers in Saudi Arabia's higher learning institutions stress that the LOBELA positively influences their teaching attitudes because it reflects English language aims (Hazaea & Tayeb, 2018). Therefore, lecturers presume that by teaching content relating to the LOBELA, they are also assisting their students in accumulating substantial knowledge on grammar rules and structure, which may be critical in the students' future endeavors such as study abroad placement. In the same regard, this exam motivates teaching staff to enhance their approaches to teaching English while preparing more learning materials to improve student performance. The test also inspires teachers to align their teaching methods with learning outcomes. Lastly, the LOBELA's influence on Saudi Arabia's higher learning was assessed through its content assessment effects. In this context, it was discovered that the LOBELA and content assessments complement each other. The researchers found that the test influenced the number of lessons provided and preparation of learning materials. In some cases, teachers limited their lessons to address only the test (Hazaea & Tayeb, 2018). Generally, the LOBELA, just like other high-stakes exams, significantly affects tutoring and learning experiences in Saudi Arabia's higher learning institutions.

### 3. History of Washback

The concept of washback is connected to the origin of public exams, which can be traced to China's entry and civil service exams. These exams served two significant purposes: choosing candidates for admission into government offices and avoiding corruption. The idea of washback was introduced in the early 1990s. In the past, scholars in applied linguistics used numerous terminologies to elucidate the idea of exam effects: "measurement-driven instruction" (Popham, 1987), "systematic validity" (Messick, 1989), "curriculum alignment" (Shepherd, 1990), "backwash" (Biggs, 1993), and "test impact" (Bachman & Palmer, 1996) (Beikmahdavi, 2016, p. 130). The notion of washback was conceived by Alderson and Wall in 1993 as a force that obliges test-takers and tutors to engage in particular tasks or activities due to exams. They postulated that exams may have immense control over the educational sector and larger society. Investigations on language testing show that exams control teaching and studying (Cheng, 2000; Beikmahdavi, 2016). To accomplish learning objectives, an exam's context and format must overlap with curriculum content. Research on language testing is centered on analyzing the features of a particular group of test-takers and how such information may be incorporated in designing language exams. According to Cheng (2000), since language exam scores represent multiple complex influences, they cannot be interpreted simplistically. Language exam scores are also affected by the exam's features and contexts, the test-takers' characteristics, the tactics used to complete the exams, and

the inferences that instructors wish to draw from them. What further complicates the evaluation of exam scores is the interrelationships between these factors. Some studies suggest that exams should guide curriculum (Cheng, 2000; Beikmahdavi, 2016). Even in the modern age, tests influence education and employment. Despite all criticisms leveled against them, exams continue to dominate the educational sector in most nations. Given the critical decisions attached to exams, it is evident that tests control many facets of life.

#### **4. Mechanisms of Washback**

Washback mechanisms are best described through several frameworks. Hughes' (1993) trichotomy concept focuses on three critical aspects: process, participants, and product. In this case, participants include test-takers, administrators, tutors, and publishers whose approaches and views toward their work may be affected by exams. The process involves any actions undertaken by the participants to enhance studying. These may include material development, altering teaching approaches, and syllabus design. The product consists of whatever is learned and its quality (Cheng, 2000; Beikmahdavi, 2016). Alternatively, Alderson and Wall (1993) concentrate on the micro factors of studying and teaching, which may be affected by tests. They proposed 15 hypotheses to indicate the studying and teaching areas that are most influenced by washback. These hypotheses emphasize that an exam will impact the depth and extent of studying and the arrangement and rate of tutoring, as well as teaching and learning (Al Hinai & Al Jardani, 2020; Cheng, 2000; Beikmahdavi, 2016). This argument is grounded in the significance of identifying the various dependent variables involved in washback to understand their relationships. It concentrates on the context, extent, sequence, methodologies, rate, and depth of studying and teaching. Alderson and Wall (1993) also suggested the need for additional studies in change and innovation in education systems as well as on performance and motivation.

Bailey's (1996) model is a combination of Hughes' trichotomy and Alderson and Wall's 15 hypotheses. It is centered on the interrelationships of the elements involved in washback. Bailey separates test-takers from instructors and studying from teaching. She also includes the role of researchers in determining the washback of exams. This framework suggests that an exam affects test-takers' and teachers' perceptions, affecting their behaviors (Cheng, 2000; Rea-Dickins & Scott, 2007). Since the conception of washback frameworks by Hughes, Bailey, and Alderson and Wall, many studies have continued to investigate washback. Watanabe's (2004) model examines the length of washback's impact. It conceptualizes washback as consisting of five dimensions: intentionality, specificity, value, intensity, and length. Intentionality examines the primary objectives behind a test. An exam may result in unintended or intended washback according to the aims of those who design or implement it. Specificity involves how wide or limited the range of an exam is (specific or general). In this case, specific washback may denote a test that examines only one particular learning factor. Value determines the positivity or negativity of the washback of an exam according to the educational context. Regarding intensity, the greater the stakes of an exam, the more value is attached to it, thus increasing the strength of washback (Ahmmed & Rahman, 2019). Alternatively, low-stakes exams have weak washback. Therefore, if an exam affects most stakeholders in an educational context, including students, learners, and other stakeholders, its washback will be strong (Cheng, 2000). Concerning length, washback can have long-term or short-term impacts on educational stakeholders, particularly students. A short-term effect occurs if test-takers maximize their learning efforts and adopt specific learning strategies during exam preparation but abandon the efforts after the exam. However, if the influence extends beyond the exam, it is considered a long-term effect (Ahmmed & Rahman, 2019).

#### **5. Validity of Washback**

Guaranteeing validity is critical when designing an exam. An exam is perceived as valid if it effectively measures what it aims to measure. Some scholars suggest that an exam's validity should be determined by its degree of positive or negative washback (Al Hinai & Al Jardani, 2020; Cheng, 2000). Shahomy et al. (2016) stress that washback validity is determined by construct validity, which involves factors like test use, effects of the exams on the learners and instructors, and the evaluation of exam scores by critical decision makers. It is essential to incorporate these exam use factors in construct validation since exams are interconnected with additional variables that interact within the learning process. Values and social meaning are critical in test validity. Messick (1989) stressed that social values are essential in determining an exam's intended or undesired outcomes. Assessing the consequential validity (social consequences) for test-takers is vital in washback studies to evaluate if exam scores meet the exam designers' purposes (Shirzadi & Amerian, 2020; Beikmahdavi, 2016). However, exam validation needs to be a continuing process to address new issues as they arise, particularly in the language testing area.

## 6. Types of Washback

Based on existing studies, washback can either positively or negatively affect the learning process and teaching. Positive washback involves test outcomes that present favorable changes in the teaching and learning process (Syafrizal & Pahamzah, 2020; Beikmahdavi, 2016). Good tests can be structured and applied as efficient teaching-learning tasks and activities to promote a positive teaching-learning process. In this case, washback can be considered positive if it reflects the course objectives' learning ideologies. For instance, the inclusion of an oral proficiency test in a class may significantly promote the training of speaking skills (Beikmahdavi, 2016). Alternatively, negative washback occurs when a test results in undesired changes and hinders students from embracing a deeper study approach, thus hampering curriculum objectives. The effects of exams can be analyzed through the choices that tutors and students make. In the case of negative washback, teachers may tutor directly for a specific exam. This implies that the exam's content will be narrowly based on the curriculum instead of covering the entire course objectives. For example, suppose a teacher only incorporates multiple-choice items to assess writing skills; the students may concentrate more on practicing such items instead of concentrating on writing skills. Thus, the discrepancy between the setting and format of the exam and the instructional management may derail the curriculum objectives because it results in the abandonment of course goals in favor of exam preparation (Thaidan, 2015; Beikmahdavi, 2016).

In most cases, instructors perceive that learners' success or failure will be reflected on the teachers, which may trigger them to exert more pressure to teach content related to an exam. In this case, inexperienced teachers may omit teaching essential writing and listening skills even though the curriculum contains these skills. This only equips students with test-taking skills rather than language skills. If an exam is regarded as high stakes, its preparation may dominate all educational processes and activities. However, if the exam contents and testing techniques vary with the course goals, it will lead to detrimental washback. Negative washback is not only harmful to the test-takers but also to tutors because it decreases their capacity to effectively teach the course contents and utilize teaching methods and materials that are congruent with efficient testing tools (Cheng, 2000; Beikmahdavi, 2016). Although negative washback adversely acts as an impediment to accomplishing educational goals, Thaidan (2015) stresses that linking curriculum objectives with test specifications can remedy such effects. Some studies propose that washback serves an intended and directed purpose (Rea-Dickins & Scott; Shirzadi & Amerian, 2020). This implies that exams are aimed at improving teaching and studying processes (positive washback). However, some scholars suggest that washback may be independent of exam quality and may be affected by other elements (Shirzadi & Amerian, 2020; Cheng, 2000). Such factors may include practicality, transparency, prestige, monopoly, accuracy, utility, and anxiety. The authors also assert that different test types may result in distinct forms of washback. In this case, the washback generated by the multiple-choice item format is distinct from the washback resulting from the open-ended response format of reading exam items.

## 7. Impacts of Washback

The impact of any test can be examined at the micro and macro levels. This helps determine the influence testing has on the education system, individual practices, policymakers, and other stakeholders (Thaidan, 2015; Beikmahdavi, 2016).

### 7.1 Impacts of Washback in Promoting Learning and Teaching

One area in which washback has been noted is the education system (Spratt, 2005). This coincides with the internalization of education, which has seen English language proficiency testing become a critical aspect for study abroad placement and global immigration (Razavipour et al., 2020). Spratt (2005), Rea-Dickins and Scott (2007), Shohamy et al. (2016), and Al Hinai and Al Jardani (2020) explore areas affected by washback in the classroom. The authors show the consequences of tests on the curriculum, such as setting standards for language testing. Studies of washback have used high-stakes tests in research conducted in Sri Lanka, Israel, and online, yielding differing results. In this case, the more significant a test is, the higher the consequence of the test on testers and test-takers. However, in some cases, students demonstrate no interest in taking tests (Razavipour et al., 2020). For example, language proficiency tests, such as the LOBELA, the TOEFL, and the IELTS, have strong washback. These tests are used to make critical decisions about test-takers. In Saudi Arabia, the LOBELA is a high-stakes exam with a significant influence on teachers and students. Thus, students are constantly under pressure to pass the test because it directly affects their future careers and job opportunities. Similarly, teachers are also obliged to design their methods and learning materials to enable students to overcome the challenges of the test. Therefore, teachers' instruction methods, content assessments, and motivations and attitudes are subject

to this test's requirements. As a result, lecturers often endeavor to meet students' expectations by preparing them to pass a high-stakes exam like the LOBELA (Hazaea & Tayeb, 2018).

The results of washback are varied. Testers often focus on changing the curriculum to satisfy test-takers' needs. Al Hinai and Al Jardani (2020) highlight an instance in which more emphasis is directed on teaching the main areas to help students garner better marks. Equally, the side effects of testing can lead to a narrowing of the curriculum to focus only on those areas that are highly examinable and disregard the remaining course content (Spratt, 2005). Moreover, when altering the content of teaching in terms of depth or intensity, teachers also need to adjust the length of time allocated for each class. In this context, extra time is assigned to exam-oriented classes, especially for high-stakes tests such as the TOEFL. Similarly, Spratt's (2005) findings also cite Andrew et al.'s study (2002), which discovered that language teachers spend two-thirds of class time working on published test-based content. In other cases, testers introduced supplementary teaching materials to help test-takers improve their skills in areas in which they do not achieve the best grades (Shohamy et al., 2016).

Tests have a considerable impact on the career and life opportunities of test-takers. These include an individual's access to educational and job opportunities. Studies have reviewed washback's effects on teachers and students, as well as on the assessments. According to Spratt (2005), the assumption is that teachers are more likely to be influenced by the fact that students are planning to take a given test. As a result, they may adapt their teaching styles and teaching materials to align with the test's requirements. For instance, a teacher may decide to teach their students near an exam period to guarantee that all the course content to be tested is covered (Al Hinai & Al Jardani, 2020). Other teachers may introduce exam-related materials like past papers or mock tests to enable test-takers to familiarize themselves with upcoming exams (Razavipour et al., 2020; Syafrizal & Pahamzah, 2020). Washback also seems to determine how and what students read. One study suggests that there was positive washback as more learners studied more frequently and formulated better study methods, like forming study groups, as a reaction to exams (Al Hinai & Al Jardani, 2020). Furthermore, many studies have also explored washback affecting teachers' feelings and attitudes (Spratt, 2005). Most teachers believe that their students' success or failure in a test directly reflects on them. Shohamy et al. (2016) observe that instructors always experience high anxiety before and after administering exams. Teachers often criticize the time pressure connected to covering course content as exams approach. They suspect that students may perform poorly in the exams and that they may be held liable for unsatisfactory results. Spratt (2005) and Cheng (2000) also discuss washback's effects on students' attitudes and feelings. For instance, test-takers tend to experience mixed feelings toward tests. Some students perceive their tests, like English exams, to be of high importance, and as a result, they work extra hard to achieve better scores. Conversely, students may perceive their Arabic exams as less important and may be reluctant to study for them (Spratt, 2005).

### *7.2 Impacts of Washback in Promoting Policy Development in Education*

Recent developments have increased the awareness that testing may have consequences beyond the classroom. According to Thaidan (2015), a test's influence can also be explored at the macro level. In further support of this finding, another study demonstrates that washback can exert significant political influence in decision-making on education (Al Hinai & Al Jardani, 2020). Based on test objectives, policymakers use their power to control the education system and curricula by imposing new teaching materials and methods. Test-takers' performance on national exams can have severe consequences for various stakeholders, such as students, teachers, and program managers. Shohamy et al. (2016) indicate that crucial decisions are arrived at based on test results. Policymakers use their authority and powers to influence those affected by tests and control the education system. This may entail reviewing program length, structuring curriculum development, and reviewing entry requirements, program length, and program delivery plans. This mechanism of power and control may extend to excluding some teachers from teaching students certain skills or terminating their services. Students are also impacted because tests are used to instill discipline in students. Decision makers also use test results, especially for classification, surveillance, and judgment (Shohamy et al., 2016; Al Hinai & Al Jardani, 2020). Thus, exam results reflect significantly on the careers and life opportunities of test-takers.

## **8. Conclusion**

Exam use in the public domain can be traced back to years when it only served as a ticket to enter public service. In other cases, tests were used to enhance accountability and transparency among civil servants. Various studies have investigated models of analysis to show how testing works on students and teachers. They include Hughes' trichotomy, the Bailey model, the Watanabe model, and Alderson and Wall's 15 hypotheses. The authors indicate that washback can either be positive (when it encourages efficient teaching and learning) or negative (when the teaching and learning process leads to undesirable outcomes). In other words, washback has a

considerable impact on both teachers and test-takers. This is because better test scores enable testers to help test-takers meet their expectations. Equally, poor student performance may lead to anxiety and negative impacts for teachers and students. In language instruction, the effects of testing play vital roles in career development and available job opportunities.

### 9. Suggestions for the Future Research

This review highlights how washback can influence the teaching and learning processes. Using the LOBELA as an example, it is evident that washback plays a vital role in teachers' and learners' behaviors and attitudes. Washback extends to the policymaking process, where it influences curriculum and educational development. The studies referenced in this review show how the impact of a test can exist at both the micro and macro levels. However, washback's effects on learning are an under-researched area. Future studies should focus on identifying types of washback at different points in time because various studies have shown that a test's impacts are highly based on the purpose of the test, the nature of the test, and other complex features (Al-Hinai & Al-Jardani, 2019; Beikmahdavi, 2016; Cheng, 2000).

### References

- Ahmmmed, R., & Rahman, A. (2019). Review of empirical studies of washback in language testing. *BMJ*, 3(1), 150-162.
- Al-Hinai, M., & Al-Jardani, K. (2020). Washback in language testing: An exploration with a focus on a specific EFL context in Oman. *International Journal of English Linguistics*, 11(1), 68-75. <https://doi.org/10.5539/ijel.v11n1p68>
- Andrews, S., Fullilove, J., & Wong, Y. (2002). Targeting washback—a case-study. *System*, 30(2), 207-223. [https://doi.org/10.1016/S0346-251X\(02\)00005-2](https://doi.org/10.1016/S0346-251X(02)00005-2)
- Beikmahdavi, N. (2016). Washback in language testing: Review of related literature first. *International Journal of Modern Language Teaching and Learning*, 1(4), 130-136.
- Cheng, L. (2000). Washback or backwash: A review of the impact of testing on teaching and learning. *Education Resources Information Center*, 1(4), 1-33.
- Hazaea, A. N., & Tayeb, Y. A. (2018). Washback effect of LOBELA on EFL teaching at preparatory program of Saudi University. *International Journal of Humanities and Applied Social Science*, 3(3), 1-10.
- Razavipour, K., Hoseini, F., & Validi, M. (2020). Washback to classroom climate: The case of an IELTS speaking preparation course. *Journal of Research in Applied Linguistics*, 11(1), 21-43.
- Rea-Dickins, P., & Scott, C. (2007). Washback from language tests on teaching, learning and policy: Evidence from diverse settings. *Assessment in Education: Principles, Policy & Practice*, 14(1), 1-7. <https://doi.org/10.1080/09695940701272682>
- Shirzadi, D., & Amerian, M. (2020). Washback effects of multiple-choice, cloze and metalinguistic tests on EFL students writing. *Indonesian Journal of Applied Linguistics*, 9(3), 536-544. <https://doi.org/10.17509/ijal.v9i3.23203>
- Shohamy, E., Donitsa-Schmidt, S., & Ferman, I. (1996). Test impact revisited: Washback effect over time. *Tel Aviv University*, 13(3), 298-317. <https://doi.org/10.1177/026553229601300305>
- Spratt, M. (2005). Washback and the classroom: The implications for teaching and learning of studies of washback from exams. *Language Teaching Research*, 9(1), 5-29. <https://doi.org/10.1191/1362168805lr152oa>
- Syafrizal, S., & Pahamzah, J. (2020). Language assessment in English language teaching: A washback of Indonesian students' test in COVID 19 situation. *Journal of Southwest Jiaotong University*, 55(4). <https://doi.org/10.35741/issn.0258-2724.55.4.40>
- Thaidan, R. (2015). Washback in language testing. *Education Journal*, 4(1), 5-8. <https://doi.org/10.11648/j.edu.20150401.12>

### Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).