Washback in Language Testing: A Literature Review of Positive and Negative Influences on Teaching and Learning

Merry Christiana

Universitas Sanata Dharma

Corespondence: merrychristiana824@gmail.com

Abstract

Washback, the influence of testing on teaching and learning, is a critical issue in language education with the potential to produce both beneficial and harmful effects. This literature review aims to provide an updated understanding of this phenomenon by synthesizing empirical and theoretical research published between 2018 and 2025. Employing a qualitative document analysis of 12 peer-reviewed articles, this study interprets recent findings through the theoretical lenses of Bachman and Palmer's (1996) model of test usefulness and Carr's (2011) framework of tests as decision-making tools. The review identifies three primary conclusions: (1) assessment design is a key determinant of washback, with product-oriented tests frequently leading to negative outcomes while process-oriented and formative tests foster positive ones; (2) washback is deeply embedded in its context, with socio-political and economic factors often overriding the test's intended effects; and (3) stakeholders, particularly teachers, are the ultimate mediators of washback, with their assessment literacy and beliefs playing a crucial role. The study concludes that washback is a complex ecosystem resulting from the interplay between the test, the context, and the people involved. Creating positive washback requires a holistic approach that improves not only test design but also invests in teacher training and mitigates negative contextual pressures.

Keyword: Washback; Language Testing; Teaching and Learning

INTRODUCTION

In educational systems worldwide, high-stakes testing has become a powerful instrument influencing policy and practice. These examinations, from national university entrance tests to international proficiency benchmarks, serve as gatekeepers to academic and professional opportunities, creating significant consequences for students, teachers, and institutions alike. The influence that these tests exert on teaching and learning is known as washback. This effect can be so profound that assessments often dictate what is taught, how it is taught, and what is learned, effectively shaping the curriculum from the end of the learning cycle backward. Given the critical role of testing, understanding the mechanisms of washback is essential for educators and policymakers who aim to design assessments that not only measure learning but also actively enhance it.

Washback is broadly defined as the influence of testing on teaching and learning. However, research has revealed this to be a highly complex, multidimensional, and often elusive phenomenon rather than a simple cause-and-effect relationship (Rahman et al., 2023). A key distinction is made between washback, which refers to effects at the micro-level of the classroom, and impact, which describes broader consequences at the macro-level of the educational system and society (Schissel, 2018; Rathnayake, 2024). Furthermore, this influence can be either positive or negative. Positive washback occurs when a well-designed test encourages beneficial teaching practices, motivates students, and aligns with curriculum goals (Alqahtani, 2021). Conversely, negative washback emerges when a test leads to undesirable consequences, such as a narrowing of the curriculum to focus only on tested skills, an overemphasis on test-taking strategies, or increased anxiety among students and teachers (Kuang, 2020).

Empirical studies from the last decade have documented this duality across diverse international contexts. Negative washback remains a persistent issue, particularly in systems dominated by high-stakes, product-oriented exams (Demirkol Orak, 2025). Research in Bangladesh, Nepal, and Türkiye has shown how national exams can compel teachers to focus narrowly on grammar and vocabulary, neglecting communicative skills like speaking and writing that are not assessed (Ali & Hamid, 2020; Dawadi, 2021; Demirkol Orak, 2025). This often results from a combination of factors, including pressure for high scores, which are treated as a form of "symbolic capital" in society, and the proliferation of a "hidden syllabus" of commercial test-preparation materials that further narrows instruction (Ali & Hamid, 2020). However, a growing body of research also provides clear evidence of positive washback. Studies have shown that when tests are perceived as valid and are well-aligned with communicative goals, they can motivate positive changes. In Thailand, an exit exam co-designed by teachers with high assessment literacy led to beneficial teaching that balanced test preparation with the development of communicative competence (Athiworakun & Adunyarittigun, 2022). Similarly, a study on progress tests in Indonesia found exclusively positive outcomes, as the formative nature of the tests motivated students, encouraged collaborative study, and improved the overall classroom atmosphere (Suryanto, 2025).

This evidence confirms that the nature of washback is not determined by the test alone but is mediated by a complex web of variables. Earlier reviews, such as Ahmmed and Rahman (2019), synthesized studies up to 2018 and concluded that washback effects were mixed and context-dependent. More recent research has moved further, arguing that to understand washback, one must look beyond the test itself to the powerful test-external factors at play (Ali & Hamid, 2020; Dawadi, 2021). These mediating factors include the socio-political context, such as political pressure for high pass rates (Ali & Hamid, 2020); economic factors, like parents' ability to afford extra resources (Dawadi, 2021); parental and family factors, including parents' educational background and the social prestige associated with test scores (Dawadi, 2021); and teacher factors, such as assessment literacy and beliefs about teaching (Rahman et al., 2023). While the importance of these factors is now widely recognized, a contemporary synthesis is needed to understand how they interact in light of the most recent empirical findings.

Therefore, this review aims to fill this gap by synthesizing empirical research on washback published between 2018 and 2025. It moves beyond a simple positive-versus-negative classification to explore the underlying mechanisms and mediating factors that shape the nature of washback in various educational contexts. To structure this analysis, this paper draws on two seminal theoretical frameworks that remain central to the field of language assessment. The framework of Bachman and Palmer (1996) is utilized for its enduring concept of test usefulness, which provides a robust lens for evaluating whether an assessment's consequences are beneficial or detrimental to learning. Complementing this, Carr's (2011) model is employed for its focus on tests as decision-making tools, which helps in analyzing how assessments influence the pedagogical choices of teachers and institutions. Although these frameworks predate the empirical studies under review, their foundational principles on test validity, consequences, and utility are timeless. Applying these established theories to the most current research allows for a rigorous and consistent analysis, providing an updated understanding of how assessments can be better designed and implemented to foster meaningful and effective learning.

METHOD

This study employs a qualitative document analysis approach. According to Bowen (2009), this is a systematic process of reviewing, evaluating, and synthesizing scholarly literature to identify themes and patterns. The research was conducted in several distinct phases to ensure a comprehensive and rigorous review.

First, the scope of the review was established, focusing on scholarly works on language testing washback published between 2018 and 2025. This timeframe was chosen to build upon existing literature reviews and incorporate the most current empirical and theoretical findings in the field. Second, a systematic data collection process was then initiated, searching academic databases such as Google Scholar, ERIC, and Scopus using keywords including "washback in language testing," "assessment backwash," "test impact on teaching," and "positive and negative washback." Third, to be included in the review, articles had to be peer-reviewed, published in English, and focus directly on the influence of language assessments on stakeholders and educational practices. This process yielded a final corpus of 12 relevant articles for analysis.

Finally, the selected documents were analyzed and synthesized. The researcher employed an inductive approach, allowing key themes to emerge directly from a detailed reading of the literature rather than imposing a pre-existing framework (Thomas, 2006). The recurrent themes identified through this process included: (a) the significant role of socio-political and contextual factors in shaping washback effects; (b) the mediating influence of teacher and stakeholder agency, beliefs, and literacy; and (c) the impact of assessment design, particularly the distinction between product-oriented and process-oriented testing. The theoretical frameworks of Bachman and Palmer (1996) on test usefulness and Carr (2011) on the decision-making function of tests were then used as analytical lenses to structure and interpret these emergent themes.

RESULT AND DISCUSSION

The findings from the reviewed literature show that washback is a complex issue shaped by many different factors. To understand how it works, this discussion is organized into three main themes that emerged from the articles: the impact of assessment design, the power of outside pressures, and the central role of the people involved.

How Test Design Shapes Washback

One of the clearest findings is that the design of the test itself plays a major role in determining whether its washback is helpful or harmful (Demirkol Orak, 2025; Rathnayake, 2024). The literature makes a useful distinction between "product-oriented" and "process-oriented" assessments. Product-oriented exams are typically recognition-based, multiple-choice tests that focus on discrete skills like grammar and vocabulary (Demirkol Orak, 2025). These tests are often used in high-stakes situations because they are easy to administer and score for large populations (Demirkol Orak, 2025). However, the evidence shows that this design frequently leads to harmful, or negative, washback.

For example, the national English exams in Türkiye (YDS and YÖKDİL) are heavily criticized for being product-oriented. Because the tests do not include a speaking section, language courses do not focus on speaking tasks, and students learn grammar rules but cannot use the language in real life (Demirkol Orak, 2025). A similar situation was found in Bangladesh, where the predictable format of the national English test, which omits listening and speaking skills, encourages teachers to narrow the curriculum and "teach to the test" (Ali & Hamid, 2020). In Nepal, the high-stakes SEE exam led students to focus almost entirely on memorization as a learning strategy, a practice that is reinforced by a test design that doesn't require active communication skills (Dawadi, 2021).

In contrast, the literature shows that process-oriented or well-aligned tests can create positive washback. These are assessments that use authentic, real-life tasks and are designed to reflect the actual goals of the curriculum. The case of the SWU-SET exit exam in Thailand is a powerful example. Because the test was designed based on communicative principles (the CEFR) and teachers understood and believed in its design, it produced positive effects. Teachers were able to prepare students for the test while also developing their broader communicative competence (Athiworakun & Adunyarittigun, 2022).

Even more striking is the evidence on formative assessments. A study on an English progress test in a high school in Indonesia found exclusively positive washback. The test motivated students to study harder, encouraged them to form study groups, and helped them develop better learning strategies (Suryanto, 2025). Because the progress test was a regular, lower-stakes part of the learning process, it became a tool for learning rather than just a final judgment (Suryanto, 2025). This shows that the design of an assessment—whether it focuses on a final product or supports the ongoing process of learning—is a critical factor in shaping its washback (Demirkol Orak, 2025; Suryanto, 2025).

The Influence of Outside Pressures

The reviewed articles strongly indicate that washback cannot be understood by looking at the test in isolation. The findings reveal that broader social, political, and economic forces are often powerful drivers that shape how a test impacts teaching and learning (Ali & Hamid, 2020; Dawadi, 2021). These "test-external" factors can create such intense pressure that they override the intended goals of a curriculum or a well-designed test.

One of the most powerful forces is the social value placed on test scores. In contexts like Bangladesh and

Nepal, test results are not just a measure of learning; they function as a form of "symbolic capital" (Ali & Hamid, 2020). High grades are directly linked to social prestige for the family, a school's reputation, and future educational and employment opportunities (Dawadi, 2021). This creates immense pressure on students, parents, and teachers to achieve high scores at any cost, which often leads to negative washback practices like teaching to the test and focusing only on exam content (Ali & Hamid, 2020). This social pressure is sometimes magnified by political goals, where governments may have an incentive to show high national pass rates as a sign of progress, which can lead to predictable and less challenging tests (Ali & Hamid, 2020).

Economic factors also play a critical role. The study in Nepal provides clear evidence of how a family's economic status directly affects washback (Dawadi, 2021). Parents with more financial resources were able to provide their children with extra learning materials and support, which influenced their study strategies and motivation. In contrast, students from poorer backgrounds lacked these resources, putting them at a disadvantage (Dawadi, 2021). This economic pressure also fuels the "hidden syllabus"—a large industry of commercial guidebooks and test-preparation materials that teachers and students rely on to secure good grades, further narrowing the focus of education (Ali & Hamid, 2020).

However, these contextual pressures are not universal. In a fascinating contrast, the case study from Oman found a lack of significant social and parental pressure on students, possibly due to the parents' different educational backgrounds (Al Hinai & Al Jardani, 2021). The absence of this intense social pressure created a different learning environment. This highlights that the nature of washback is highly dependent on the specific context in which the test operates. These findings strongly support the view that washback is deeply embedded in its environment. Therefore, any effort to promote positive washback must address these powerful external factors, not just the features of the test itself.

The Role of Teachers and Students

Beyond test design and outside pressures, the reviewed articles show that the people involved—primarily teachers and students—are the final and most important actors in shaping washback. Their personal skills, beliefs, and actions often determine the ultimate effect of a test.

Teachers are consistently identified as the key agents in creating positive washback in the classroom (Rahman et al., 2023; Schissel, 2018). One of the most important factors is teacher assessment literacy. The study in Thailand showed that when teachers are knowledgeable about assessment and are part of the test development process, they can create positive washback. Because they understood the goals of the test, they could align their teaching in a way that helped students pass the exam while also building real communication skills (Athiworakun & Adunyarittigun, 2022). Teachers' beliefs about a test are also critical. In Saudi Arabia, for example, teachers saw "teaching to the test" as a good thing because they believed the test was high-quality and matched their teaching goals (Alqahtani, 2021). This shows that if teachers believe in a test, their teaching is more likely to be positive and motivating.

The role of students is just as important. Their attitudes and learning habits can strongly influence washback. The study in Oman showed how students who are not very motivated and only focused on passing the exam can contribute to a negative washback environment (Al Hinai & Al Jardani, 2021). Similarly, the Nepal study found that students' habit of memorizing answers was a major part of their learning culture, and this strategy was used to prepare for the test regardless of its specific goals (Dawadi, 2021). However, students can also drive positive change. In the Indonesian study, students responded to a progress test by becoming more motivated and forming study groups, showing that a good test can encourage positive actions from students (Suryanto, 2025).

Ultimately, the evidence shows a dynamic relationship between the test, the context, and the people involved. A test does not create washback on its own. Instead, positive washback is something that is actively created by well-informed teachers and engaged students.

CONCLUSION

This literature review aimed to provide an updated understanding of washback by synthesizing empirical and theoretical research published between 2018 and 2025. The analysis of these recent studies confirms that washback is not a simple, direct result of a test. Instead, its nature is shaped by a complex interplay of factors.

The findings revealed three overarching themes: first, that the design of an assessment—specifically, whether it is product-oriented or process-oriented—is a critical factor in determining the direction of washback. Second, that washback is deeply embedded in its socio-political and economic context, with outside pressures often being more influential than the test itself. Third, that teachers and students are the ultimate agents who mediate washback through their beliefs, skills, and actions.

Based on these findings, this review concludes that the traditional "test-causes-effect" model is insufficient to explain the washback phenomenon. The primary conclusion of this study is that washback should be understood as a dynamic ecosystem created by the interaction between the assessment, the context, and the people involved. Positive washback is not something a test can create on its own; it must be actively fostered by knowledgeable stakeholders within a supportive educational and social environment. A change in any one of these elements can alter the entire system, meaning that simply introducing a "good" test into an unsupportive context is unlikely to produce the desired positive changes.

These conclusions lead to several practical recommendations for policymakers, test designers, and educators seeking to promote positive washback. First, there should be a conscious shift in assessment design away from a sole reliance on product-oriented, multiple-choice exams. Integrating more process-oriented and formative assessments, such as the progress tests found to be successful in Indonesia, can encourage deeper, more engaged learning. Second, policymakers must work to reduce the intense negative pressures associated with high-stakes tests. Providing students with multiple pathways to meet requirements, such as taking a remedial course or submitting alternative test scores as practiced in Thailand, can lower the stakes and reduce the anxiety that leads to negative teaching practices. Finally, the most critical investment is in teacher assessment literacy. Since teachers are the key agents in the classroom, providing them with high-quality professional development is essential for creating positive washback.

This review also highlights several directions for future research. Building on recent calls in the literature, future studies should focus less on merely identifying washback effects and more on understanding their underlying causes. There is a need for research that explores the cognitive experiences of stakeholders to understand their beliefs and decision-making processes. Furthermore, future work should address issues of equity and fairness, examining how washback affects students from different socio-economic backgrounds. By continuing to explore these complex dynamics, the field can move closer to designing and implementing assessments that not only measure learning but consistently improve it for all students.

References

Ahmmed, R., & Rahman, K. A. (2019). Review of empirical studies of washback in language testing. *Bangladesh Maritime Journal, 3*(1), 150–162.

Al Hinai, M. K., & Al Jardani, K. S. (2021). 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. doi:10.5539/ijel.v11n1p68

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. doi:10.1080/15434303.2020.1717495

Alqahtani, F. (2021). The impact of language testing washback in promoting teaching and learning processes: A theoretical review. *English Language Teaching, 14*(7), 21–27. doi:10.5539/elt.v14n7p21

Athiworakun, C., & Adunyarittigun, D. (2022). Investigating washback effects on teaching: A case study of an exit examination at the higher education level. *LEARN Journal: Language Education and Acquisition Research Network, 15*(2), 776–801.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests.* Oxford, UK: Oxford University Press.

Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal, 9*(2), 27–40.

Carr, N. T. (2011). *Designing and analyzing language tests*. Oxford, UK: Oxford University Press.

Dawadi, S. (2021). Factors influencing washback of a high-stakes English-as-a-foreign-language test. *TESL-EJ, 25*(3), 1–16.

Demirkol Orak, S. (2025). Washback effect and English language testing in Türkiye: A critical review study. *New Era International Journal of Interdisciplinary Social Researches, 10*(28), 155–168.

Kuang, Q. (2020). A review of the washback of English language tests on classroom teaching. *English Language Teaching, 13*(9), 10–18. doi:10.5539/elt.v13n9p10

Rahman, K. A., Rukanuddin, M., Chowdhury, M. S. Y., Ahmed, S., & Seraj, P. M. I. (2023). Recognizing stakeholders and factors mediating washback in language testing. *Education Research International, 2023*, Article ID 5548723. doi:10.1155/2023/5548723

Rathnayake, W. M. P. Y. B. (2024). From tests to teaching: Exploring the washback effect in language pedagogy. *The Journal of Desk Research Review and Analysis, 2*(2), 145–158.

Schissel, J. L. (2018). Test impact and washback. In J. I. Liontas (Ed.), *The TESOL encyclopedia of English language teaching* (pp. 1–6). Hoboken, NJ: John Wiley & Sons, Inc. doi:10.1002/9781118784235.eelt0353

Suryanto, J. (2025). Washback effect: How it works on EFL students. *Celtic: A Journal of Culture, English Language Teaching, Literature and Linguistics, 12*(1), 323–339.

Thomas, D. R. (2006). A general inductive approach for analyzing qualitative evaluation data. *American Journal of Evaluation, 27*(2), 237–246.