

# Washback Effects in High-Stakes Language Testing within the EFL Context: A Multidimensional Review

He Che<sup>1</sup>, Shaidatul Akma Adi Kasuma<sup>1\*</sup>

<sup>1</sup> School of Languages, Literacies and Translation, Universiti Sains Malaysia, Pulau Pinang, Malaysia

\*Corresponding Author: [shaidatul@usm.my](mailto:shaidatul@usm.my)

Received: 30 September 2025 | Accepted: 15 November 2025 | Published: 1 December 2025

DOI: <https://doi.org/10.55057/ijares.2025.7.6.37>

---

**Abstract:** *High-stakes English language assessments significantly influence EFL teaching, learning, and policy, a phenomenon known as washback. This paper reviews recent studies (2019–2024) to examine these effects on instructional practices, student strategies, assessment design, and policy reforms. The analysis reveals that test design is a pivotal determinant of washback quality; assessments integrating communicative tasks foster positive outcomes, while those focusing on rote memorization lead to negative effects like curricular narrowing. Furthermore, socio-economic disparities and institutional constraints mediate these effects. Crucially, the review concludes that isolated policy reforms often fail without systematic support, including teacher training and alternative assessment models. Future research should pursue cross-cultural longitudinal studies to explore how test design and policy interventions can collaboratively promote sustainable language learning outcomes.*

**Keywords:** Washback, high-stakes testing, EFL education, language assessment, policy reform

---

## 1. Introduction

In an increasingly globalized world, language proficiency has emerged as a crucial determinant of academic, professional, and social success (Kirkpatrick, 2015; Van Tubergen & Kalmijn, 2022). As international communication continues to expand, English has firmly established itself as the dominant lingua franca. This development has led to the widespread adoption of standardized language assessments such as TOEFL, IELTS, and national matriculation exams. These high-stakes tests function as gatekeepers that influence university admissions, scholarship opportunities, and career prospects (French, Dickerson, & Mulder, 2024). Beyond their evaluative roles, these assessments exert considerable influence on educational systems by shaping curriculum design, instructional practices, and student learning behaviors within the EFL context (Cheng, 2015).

This pervasive influence is commonly referred to as washback—the effect that tests have on teaching and learning (Alderson & Wall, 1993). While some researchers contend that high-stakes tests foster targeted learning and enhance instructional clarity (Dawadi, 2021), others emphasize their unintended consequences. These may include curriculum narrowing; an excessive focus on test-taking strategies; and reduced opportunities for authentic language use (Arthur & Partey, 2023). Furthermore, cross-cultural differences in educational policies and

institutional responses complicate the generalizability of findings regarding washback effects across diverse EFL contexts (Owusu, 2021).

Despite the extensive body of research on washback effects, several key gaps remain. First, the majority of studies focus on short-term washback, with limited longitudinal research examining its long-term impact on language acquisition and instructional quality (Nguyen, 2023). Second, while many studies document the effects of washback in individual national contexts, comparative cross-cultural research is scarce, leaving questions about how different educational systems mediate testing influence (Güloğlu-Demir & Kaplan-Keleş, 2021). Third, although alternative assessments such as performance-based testing and task-based language teaching (TBLT) have been proposed to mitigate negative washback, empirical validation of these approaches remains limited (Fasih, 2022). Addressing these gaps is essential for refining assessment practices, supporting teacher autonomy, and improving language learning outcomes.

This paper provides a comprehensive review of the literature on washback effects in high-stakes language testing within the EFL context. Specifically, it aims to (a) synthesize key findings on teacher practices, student learning strategies, assessment design, and policy impacts; (b) critically evaluate the strengths and limitations of existing research; and (c) identify areas for future inquiry. To systematically analyze these complex interactions, this review is guided by an integrated theoretical framework synthesizing the hypotheses of Alderson and Wall (1993), the trichotomy of Hughes (1996), and the multidimensional model of Watanabe (2004)."

## 2. Theoretical Framework

Washback, defined as the influence of testing on teaching and learning, has been extensively examined within the realm of language assessment research. Alderson and Wall (1993) were pioneers in washback theory, proposing 15 hypotheses that elucidate how tests shape various educational components. These hypotheses indicate that tests impact instruction, curriculum content, pedagogical approaches, student learning strategies, as well as the sequencing and depth of teaching. Moreover, the degree of washback is contingent upon the stakes associated with a test—examinations carrying significant consequences exert a more pronounced influence on both educators and learners; conversely, low-stakes assessments may yield minimal to no washback effects. It is crucial to note that washback is not uniform; individual teachers and students experience varying degrees of impact based on personal and contextual factors.

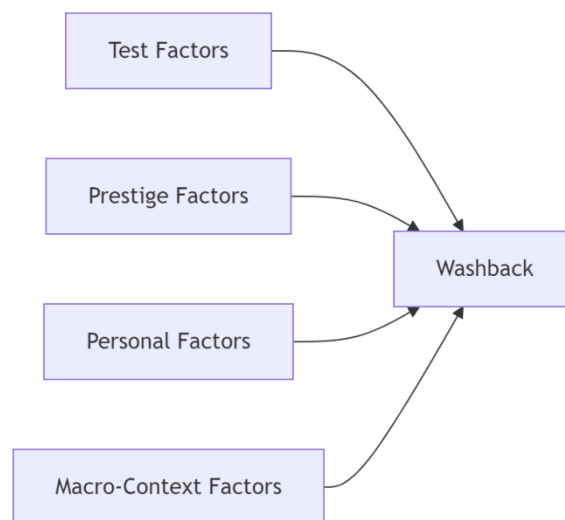
Building upon this foundational work, Hughes (1993, as cited in Bailey, 1996) developed a trichotomy model categorizing washback effects into three interrelated components: participants (e.g., teachers, students, material developers), processes (the actions undertaken in response to tests), and products (learning outcomes alongside overall educational quality). This model offers a structured framework for analyzing how tests affect educational ecosystems. Saville (2009) further refined this framework into what is known as the 3Ps model. While it similarly emphasizes people, processes, and products, it places greater emphasis on the interactions among test-takers, educators, and institutional factors in shaping washback effects.

**Table 1: Dimensions of Test Washback Impact**

Domain	Impact Manifestations
People	Teachers' methodology shifts; Students' learning strategies

Processes	Curriculum narrowing; Increased practice tests
Products	Imbalanced proficiency; Certification outcomes

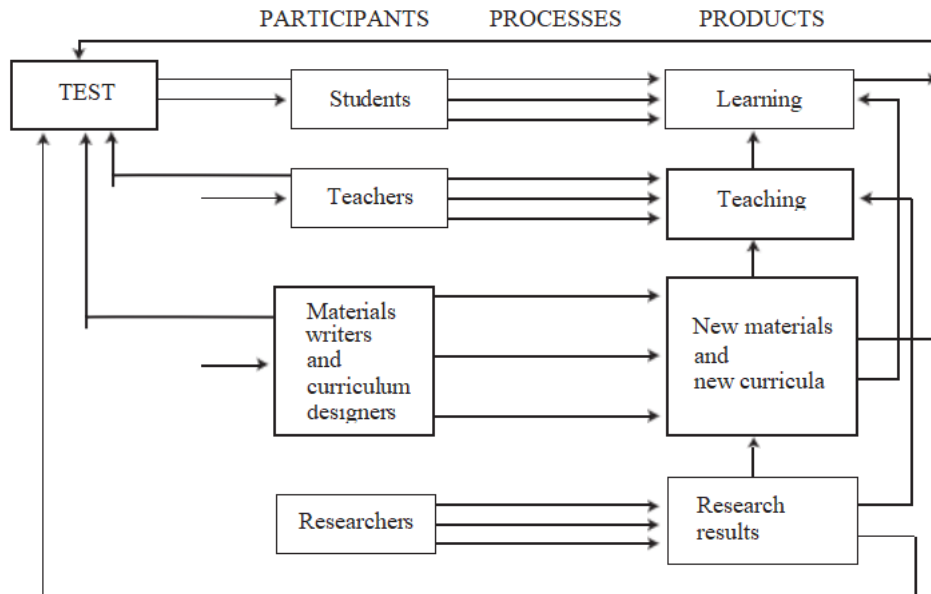
Extending these perspectives, Watanabe (2004) introduced a more detailed conceptualization of washback by identifying five key dimensions: specificity (whether the impact is broad or test-specific), intensity (degree of influence), length (short-term vs. long-term effects), intentionality (whether washback is deliberately designed into an exam), and value (positive or negative consequences). Watanabe also identified key mediating factors that shape washback effects, including test factors (exam content and format), prestige factors (the social status of a test), personal factors (students' and teachers' attitudes), and macro-context factors (educational policy, institutional constraints). These factors help explain why the same exam may produce varying effects across different educational settings.



**Figure 1: Factors Influencing Language Testing**

One of the most significant discussions in washback theory is the distinction between positive and negative washback. Negative washback occurs when tests promote narrow, test-specific learning rather than broader language proficiency. This can lead to curriculum narrowing, excessive focus on exam techniques, and avoidance of communicative language use (Alderson & Wall, 1993; Brown, 2004). For example, in contexts where exams primarily assess grammar and reading comprehension, students may neglect speaking and listening skills, resulting in imbalanced language development. Conversely, positive washback occurs when a test aligns with pedagogical best practices, encouraging effective instruction and meaningful learning experiences (Taylor, 2005). When assessment tasks closely reflect real-world language use, teachers are more likely to adopt interactive, communicative teaching methods that foster holistic language development.

These theoretical models provide a comprehensive framework for understanding how and why high-stakes testing influences EFL teaching, learning strategies, assessment design, and educational policy. This paper draws on these models to analyze empirical findings, identifying the conditions under which assessments contribute to constructive or detrimental washback effects in various EFL contexts.

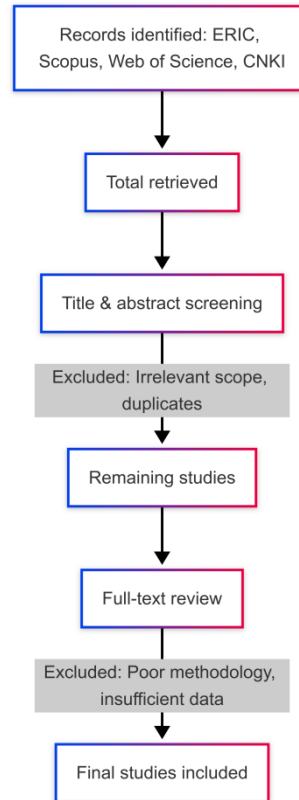


**Figure 2: Basic Model of Washback (Baily, 1996)**

### 3. Literature Search and Selection Methods

To ensure a comprehensive and systematic review, a multi-database search strategy was employed. ERIC, Scopus, Web of Science, and CNKI were searched using key terms such as “washback,” “high-stakes testing,” “EFL,” “English as a Foreign Language,” “teaching,” “learning,” “assessment,” and “policy.” The search was limited to peer-reviewed journal articles, conference papers, and high-quality dissertations published between 2019 and 2024 to ensure the inclusion of the most recent empirical findings.

The selection process involved multiple screening stages. First, titles and abstracts were reviewed to exclude irrelevant studies and duplicates. Next, full texts were examined for methodological quality, clarity, and relevance to the review’s objectives. Studies were included if they (1) focused on high-stakes language testing, (2) were conducted in EFL contexts or had significant implications for EFL education, and (3) employed empirical methods (quantitative, qualitative, or mixed methods). In total, 10 studies were included, covering washback effects on teaching, student learning, assessment design, and policy.



#### 4. Major Findings from the Studies Reviewed

This paper reviewed ten empirical studies of washback in language teaching conducted around the world between 2019 and 2024. The major finding of these studies are summarised and analysed below:

##### 4.1 Washback Effects on EFL Teaching

High-stakes language assessments exert a significant influence on English as a Foreign Language (EFL) teaching, shaping instructional methodologies, curriculum design, and classroom practices. This section reviews studies that examine how these examinations affect teachers' pedagogical decisions, often resulting in instructional narrowing or adaptations influenced by socio-economic and policy factors.

Test-driven instruction and pedagogical constraints are prevalent in high-stakes test preparation. Sağlam and Farhady (2019) found that university-level English proficiency exams in Turkey led to rigid, test-centric teaching approaches that constrained communicative instruction. Their mixed-method study—grounded in classroom observations and focus-group interviews with 14 instructors—revealed that educators prioritized test strategies such as elucidating exam formats, drilling students on predictable question types, and emphasizing memorization over interactive activities. This shift consequently diminished opportunities for authentic language use. However, the small sample size of the study and its focus on a single institution limit the generalizability of its findings. Furthermore, reliance on self-reported data raises concerns regarding potential bias in how teachers represented their instructional practices.

Similarly, Estaji and Ghiasvand (2021) examined the impact of IELTS preparation in Iran, reinforcing findings from Turkey. Their study, integrating a quantitative survey of 120 educators with in-depth interviews of 15 participants, showed that IELTS-oriented instruction led to a highly structured and formulaic approach. Teachers relied on memorized essay structures, repetitive listening drills, and fixed response patterns, which improved test performance but restricted instructional creativity. Unlike Turkish instructors, who experienced rigid instruction as an unintended consequence, Iranian teachers deliberately structured lessons to align with IELTS expectations. However, both studies faced similar methodological concerns—self-reported data and the absence of classroom observations limit verification of the reported teaching practices.

Socio-economic factors further shape EFL teaching approaches under high-stakes testing pressures. Dawadi (2021) examined how Nepal's Secondary Education Examination (SEE) influenced instructional practices in different socio-economic contexts. Through oral diaries and semi-structured interviews with students and parents, the study found that students from privileged backgrounds benefited from structured preparatory courses that encouraged strategic learning and engagement. In contrast, disadvantaged students relied more on rote memorization due to limited access to quality preparatory resources. Unlike previous studies that primarily examined teachers' classroom choices, this study highlights how socio-economic disparities indirectly affect EFL teaching, as resource availability shapes the type of instruction students receive. However, the study's reliance on interviews alone makes it difficult to establish causality between socio-economic status and instructional methodologies.

Beyond classroom-level decisions, institutional policies and exam design influence EFL teaching strategies. Homran and Asassfeh (2023) analyzed how Jordan's General Secondary Certificate English Exam (JGSCEE) shaped teacher instruction. Their large-scale survey found that, as in Turkey and Iran, Jordanian teachers emphasized exam preparation over communicative competence. However, while Nepalese teachers adjusted instruction based on socio-economic constraints, Jordanian teachers' strategies were largely dictated by institutional policies and exam requirements. Although the study benefits from a large sample size, its reliance on survey-based data limits insight into how specific pedagogical choices were implemented.

These studies collectively demonstrate that EFL teaching under high-stakes testing conditions tends to shift toward test-oriented instruction, prioritizing exam techniques over communicative competence. While some studies attribute this trend to teacher-driven instructional choices (Sağlam & Farhady, 2019; Estaji & Ghiasvand, 2021), others highlight the role of socio-economic disparities (Dawadi, 2021) or policy-driven constraints (Homran & Asassfeh, 2023). Despite contextual differences, a consistent theme is the reduction in interactive and communicative teaching in favor of test-aligned instruction. Future research should adopt cross-context comparative approaches, longitudinal methodologies, and mixed-method designs to explore how instructional decisions, socio-economic conditions, and institutional policies interact to shape EFL teaching under high-stakes testing conditions.

In summary, the studies in this section corroborate the theoretical propositions of Alderson and Wall (1993) and Hughes (1996) that tests profoundly influence teaching processes and people (instructors). The findings reveal that washback is not monolithic but is mediated by macro-context factors (Watanabe, 2004), such as socio-economic constraints (Dawadi, 2021) and institutional policies (Homran & Asassfeh, 2023). This challenges a simplistic view of

washback as merely test-driven, highlighting instead a complex interplay where teacher agency, resource availability, and policy mandates collectively shape instructional outcomes.

#### **4.2 Washback Effects on EFL Students' Learning**

High-stakes language evaluations significantly influence EFL students' learning by affecting their study strategies, motivation, and general engagement. Some research indicate beneficial washback effects, while others reveal detrimental implications that impede long-term language development. This section juxtaposes essential study findings regarding both characteristics, taking into account variations in assessment forms, student views, and institutional impacts.

Certain students derive advantages from organised test preparation, utilising examinations as a means for language enhancement and motivation. Sadeghi et al. (2021) executed a survey-based investigation with 200 overseas TOEFL and IELTS candidates, indicating that numerous students perceived these examinations as a method to establish definitive academic objectives. The research indicated that systematic exam preparation fostered self-discipline, vocabulary development, and focused skill improvement, especially in writing and reading comprehension. Nevertheless, the study was deficient in qualitative classroom data, complicating the assessment of how preparation tactics differed across various educational environments.

Nguyen (2023) investigated the influence of IELTS as a graduation prerequisite for Vietnamese university students through a longitudinal survey methodology. The research indicated that several students viewed the examination as a means to methodically enhance their English proficiency, especially in writing and reading, due to the necessity of participating in extended essay composition and organised comprehension tasks. In contrast to TOEFL and IELTS applicants who engaged in voluntary preparation for high-stakes assessments (Sadeghi et al., 2021), IELTS examinees in Vietnamese universities underwent compulsory test preparation, resulting in more polarised student reactions. Some perceived the exam as advantageous, while others cultivated adverse perceptions of its fairness and applicability to practical language use. The research would have been enhanced by interviews or classroom observations to investigate the particular learning strategies employed by students who perceived the test favourably.

Despite these advantages, numerous students employ test-oriented tactics that restrict communicative proficiency, frequently emphasising rote memorisation rather than substantive language application. Owusu (2021) executed an extensive study including 344 Ghanaian EFL students preparing for the West African Senior School Certificate Examination (WASSCE). The results indicated that students predominantly focused on reading and writing abilities pertinent to tests, while disregarding speaking and listening, as these were not explicitly evaluated. The study determined that the lack of a speaking or listening component in the examination resulted in a skewed language learning methodology, promoting memorisation techniques above communicative proficiency. Nevertheless, the study employed a cross-sectional design, constraining the capacity to monitor whether students modified their learning behaviours over time.

Mahmud et al. (2021) examined alterations in the study habits of Malaysian students in preparation for the Malaysian University English Test (MUET) employing a pre-post intervention design. The research indicated that rigorous exam preparation resulted in a decrease in spontaneous speech and innovative language generation. Students utilised standardised templates, memorised word lists, and test-specific writing patterns to optimise their scores. The study's experimental methodology yielded definitive evidence of test-related

behavioural changes, supporting the notion that exams prioritising formulaic responses foster inflexible learning patterns. Nevertheless, the study did not investigate whether students could regain their communicative competencies post-test, raising enquiries over the enduring effects of test preparation methodologies.

The analysed research underscore the advantages and disadvantages of high-stakes language assessments on EFL student acquisition. Although certain kids develop drive and organised study practices, many rely on rote memorisation, which constrains long-term communication proficiency. The nature of exams and the institutional environment are pivotal—students preparing for speaking-inclusive assessments (e.g., TOEFL, IELTS) saw more favourable washback, while reading- and writing-centric tests (e.g., WASSCE, MUET) perpetuated inflexible learning processes. Nevertheless, the majority of research depend on survey data, missing longitudinal and classroom-based observations. Future research ought to integrate various methodologies to enhance comprehension of the impact of exam preparation on students' language development throughout time.

Collectively, these findings underscore the theoretical distinction between positive and negative washback. They validate Watanabe's (2004) dimension of value by demonstrating that the nature of the test itself—specifically whether it assesses communicative competence or discrete skills—is a primary determinant of learning outcomes. The studies show that washback on learners (People) manifests in their learning processes (e.g., memorization vs. strategic practice), ultimately affecting the product of their language proficiency. This reinforces the model that test design is a powerful mediator of student behavior and skill development.

#### **4.3 Washback Effects on EFL Assessment Design**

Assessment design plays a crucial role in shaping instructional practices and student learning outcomes in EFL contexts. This section reviews studies on the washback effects of major standardized English tests, focusing on how test-driven instruction influences communicative competence and curriculum design.

Ahn (2022) examined the washback effects of TOEIC at private educational institutions (hagwons) in Korea. The study employed a sequential mixed-methods approach, surveying 103 students and 22 teachers, and found that TOEIC-oriented instruction predominantly emphasised test-taking tactics over communicative competence. This test-driven methodology produced a limited curriculum, overlooking the cultivation of practical language competencies. The study did not examine whether these outcomes are confined to private schools or investigate how regional policy variations may influence washback. Moreover, the study concentrated on a singular test type, which may not comprehensively reflect the influence of many evaluation methods across distinct educational contexts. Yue (2022) and Gu (2022) both examine the CET4 and CET6 exams in China, but their approaches differ in terms of research methods. Yue's study focuses on the integration of information literacy into English language instruction, with an emphasis on its potential washback effects. Yue used a qualitative approach, analyzing existing curriculum frameworks and assessment designs to argue for the inclusion of tasks that promote critical thinking and practical language application. The study highlights the need for reforms in assessment design but lacks empirical data on the implementation of these changes and their long-term effects. Gu (2022) takes a more empirical approach, investigating the gap between English language curriculum design and its practical implementation in Chinese institutions. Gu conducted a combination of qualitative and quantitative research, including surveys and interviews with teachers and students to understand the impact of high-stakes exams on language learning. The study focuses on the

negative washback effects of CET4 and CET6, emphasizing the need for communicative teaching methods that prioritize real-world language use. However, Gu's study also falls short in examining the specific challenges faced by individual schools, such as variations in resources and teacher training. Both studies highlight the need for reforms in assessment and teaching but fail to provide comprehensive empirical evidence on the challenges of implementation across diverse educational contexts.

The studies reviewed highlight both the limitations and potential reforms of high-stakes testing in EFL education. While exams like TOEIC and CET4/6 drive test-focused instruction, they also present opportunities for curriculum innovation when assessment design aligns with communicative goals. However, challenges in policy implementation and resource allocation remain key barriers to reform.

The review of assessment design underscores a central tenet of washback theory: tests determine what is taught and how. The studies on TOEIC and CET affirm that the format and content of high-stakes exams (Test Factors in Watanabe's model) can lead to a narrowing of the curriculum and a focus on test-taking strategies, thus validating concerns about negative washback. However, the discussion of potential reforms also points to intentionality (Watanabe, 2004)—the possibility of designing assessments to deliberately foster positive washback by aligning more closely with communicative language teaching principles, thereby influencing teaching processes and learning products more constructively.

#### **4.4 Washback Effects in EFL Education Systems and Policy**

High-stakes language testing shapes national education policies by influencing curriculum decisions, teacher training, and resource allocation. While some policies aim to reduce test dependency, others struggle to implement meaningful reforms. This section examines research on the policy-level impact of high-stakes testing across different EFL contexts.

Dawadi (2021) found that socio-economic factors play a crucial role in shaping test washback, particularly in Nepal's Secondary Education Examination (SEE). Through oral diaries (n=72) and semi-structured interviews (n=24), the study revealed that social status, economic background, and parental education influenced students' test preparation. Unlike studies focusing on instructional practices, this research highlights the broader societal pressures that reinforce test-focused learning. However, the study lacks direct classroom observation, making it difficult to establish how these factors influence teachers' instructional decisions. In contrast to Nepal, where test washback is shaped by socio-economic disparities, Indonesia's policy shift eliminated national English exams to grant schools more flexibility. Ashadi et al. (2022) conducted semi-structured interviews with educators and policymakers and found that while this reform increased instructional autonomy, it also led to inequities in English proficiency due to inconsistencies in teaching quality across schools. Unlike SEE in Nepal, where high-stakes exams create uniformity in instruction, the removal of exams in Indonesia resulted in greater variation in learning outcomes, raising concerns about educational equity. However, this study does not explore alternative assessment models that could mitigate these disparities.

Nguyen (2023) examined Vietnam's English exit exam reforms, revealing that policy efforts alone cannot shift entrenched teaching practices. This five-year longitudinal study found that, despite curriculum mandates promoting communicative language teaching (CLT), most schools continued test-oriented instruction due to teacher training gaps and outdated assessment systems. Compared to Indonesia's policy shift, where eliminating national exams caused instructional variation, Vietnam's policy failed to change classroom practices, showing

that policy reforms must be supported by long-term teacher development and assessment redesign. However, this study focuses on school-level implementation and does not examine how national policy adjustments interact with local educational governance.

Beyond policy mandates, sociocultural values also shape test washback. Puspitasari (2024) explored the cultural dimensions of Indonesia's national examination through qualitative interviews with eight secondary school teachers. The study found that social pressure and collective attitudes reinforced test-oriented instruction, even after national policies attempted to shift towards communicative teaching. Unlike Nguyen's (2023) study, which emphasized structural constraints, this research suggests that deep-rooted cultural expectations also limit test reform success. However, the study's small sample size raises questions about generalizability across different Indonesian regions. These studies collectively highlight that test-driven policy reforms face challenges from socio-economic, institutional, and cultural constraints. While removing national exams (Indonesia) creates instructional disparities, maintaining them (Nepal, Vietnam) reinforces test-centric learning. The success of policy reforms depends on teacher training, alternative assessment models, and addressing cultural expectations. Future research should explore how policy shifts can balance national standardization with localized flexibility to create sustainable improvements in EFL instruction.

## 5. Discussion and Conclusion

The reviewed studies consistently highlight instructional narrowing, where teachers prioritize test-specific strategies, reducing opportunities for interactive learning (Sağlam & Farhady, 2019; Estaji & Ghasvand, 2021). However, in contexts where assessments incorporate communicative components, washback effects can be more constructive (Wiyaka, 2020; Yue, 2022). Similarly, while students in exam-driven systems tend to rely on memorization (Owusu, 2021; Mahmud et al., 2021), those preparing for tests with speaking and listening elements exhibit more structured skill development (Sadeghi et al., 2021; Nguyen, 2023). These variations suggest that assessment design plays a crucial role in shaping learning behaviors, with well-structured exams encouraging meaningful language use while rigid, test-driven models reinforce superficial learning strategies.

At the policy level, reforms aimed at reducing test dependency face challenges, as seen in Indonesia's national exam abolition, which increased instructional autonomy but led to greater learning disparities between schools (Ashadi et al., 2022). Similarly, Vietnam's curriculum reforms encouraging communicative teaching had minimal classroom impact due to entrenched instructional habits and insufficient teacher training (Nguyen, 2023). Despite policy shifts, sociocultural expectations also sustain test-focused learning, as observed in Indonesia, where standardized exams continue to shape educational practices even after formal policy changes (Puspitasari, 2024). These findings suggest that while policy adjustments may intend to reduce test reliance, the reality of implementation depends on broader institutional and cultural factors.

The relationship between washback effects and broader educational frameworks aligns with Alderson and Wall's (1993) hypotheses, confirming that high-stakes tests influence curriculum, pedagogy, and learning strategies. However, research highlights that institutional constraints and socio-economic conditions moderate these effects (Dawadi, 2021). For example, while TOEIC preparation in Korea fosters rigid test-taking skills and narrow curricula (Ahn, 2022), alternative assessments in China have been linked to broader learning objectives

and engagement (Gu, 2022). These variations underscore the importance of test design, where assessments reflecting authentic language use can encourage communicative instruction rather than mechanical test preparation. To ensure positive washback, policymakers must integrate test reforms with teacher training and curriculum restructuring, rather than relying solely on assessment changes to drive instructional shifts.

Despite these insights, methodological limitations persist. Many studies rely on self-reported data, making it difficult to verify instructional practices through direct classroom observations (Sağlam & Farhady, 2019; Estaji & Ghiasvand, 2021). Furthermore, longitudinal research remains limited, preventing a deeper understanding of how washback evolves over time (Nguyen, 2023). Additionally, existing studies are largely concentrated in Asian EFL contexts, leaving gaps in understanding washback effects in regions such as Latin America, Africa, and Europe (Owusu, 2021). Another challenge is publication bias, with studies often focusing on negative washback while fewer explore constructive test influences, limiting a balanced understanding of how assessments shape language learning.

To address these gaps, longitudinal and cross-cultural studies are necessary to track washback effects across time and different educational systems. Further research should evaluate alternative assessments, examining whether communicative and task-based testing methods effectively reduce negative washback (Wiyaka, 2020; Yue, 2022). In addition, more studies should explore policy implementation challenges, particularly the role of teacher training and resource allocation in determining the success of assessment reforms (Nguyen, 2023). Understanding how different stakeholders—teachers, students, policymakers—respond to testing changes will be critical in developing more effective and equitable language assessment systems.

### **Acknowledgement**

The authors would like to express sincere gratitude to everyone who contributed, both directly and indirectly, to the completion of this study.

### **Conflict of Interest Statement**

The authors declare that there is no conflict of interest regarding the publication of this study.

### **References**

- Ahn, J. (2022). A study of the washback effects of a standardized test on the EFL context. *The European Conference on Language Learning 2022: Official Conference Proceedings*.
- Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14(2), 115-129.
- Arthur, F., & Partey, P. (2023). Washback effects of high-stakes test: Perspectives of economics teachers and students. *Asian Journal of Education and Social Studies*.
- Ashadi, A., Margana, M., Mukminatun, S., & Utami, A. B. (2022). High-stakes testing cancellation and its impact on EFL teaching and learning: Lessons from Indonesia. *International Journal of Language Education*.
- Atmojo, A. E. P. (2021). Facilitating positive washback through authentic assessment in EFL assessment. *Acitya: Journal of Teaching and Education*.
- Bailey, K. M. (1996). *Language teacher supervision*. Cambridge University Press.
- Bailey, K. M. (1996). Working for washback: A review of the washback concept in language testing. *Language Testing*, 13(3), 257-278.
- Brown, D. H. (2004). *Language assessment principles and classroom practices*. Pearson Education.

- Cheng, L. (2015). Impact of high-stakes language tests on EFL education: A critical perspective. *Language Testing Journal*, 32(1), 53-72.
- Dawadi, S. (2021). Factors influencing washback of a high-stakes English-as-a-Foreign-Language test. *Open Research Online Journal*, 5(2), 202–213.
- Estaji, M., & Ghiasvand, F. (2021). IELTS washback effect and instructional planning: The role of IELTS-related experiences of Iranian EFL teachers. *Journal of English Language Teaching and Learning*, 13(1), 163–192.
- Fasih, P. (2022). A critical literature review on the washback effect of task-based vocabulary assessment on Iranian EFL learners. *Frontiers in Education*, 7.
- French, S., Dickerson, A., & Mulder, R. A. (2024). A review of the benefits and drawbacks of high-stakes final examinations in higher education. *Higher Education*, 88(3), 893-918.
- Güloğlu-Demir, C., & Kaplan-Keleş, Ö. (2021). The impact of high-stakes testing on the teaching and learning process. *Journal of Pedagogical Research*, 5(2), 119-137.
- Gu, J. (2022). *College English Education in China: From Curriculum Formulation to Implementation Reality* (Doctoral dissertation, Saint Louis University).
- Gu, Y. (2023). Cultural bias in IELTS listening tests and its effect on non-Western test-takers: An item analysis and test-taker feedback evaluation. *International Journal of Applied Linguistics and English Studies*.
- Homran, M. M. A., & Asassfeh, S. M. (2023). EFL high-stakes exams: Are we leading teachers as language teachers or test teachers? *Journal of World Englishes and Educational Practices*, 5(3), 34–52.
- Kirkpatrick, A. (2015). *World Englishes: Implications for international communication and English language teaching*. Cambridge University Press.
- Mahmud, N., Lateh, N., Mahmud, N., Hassan, A. A., Mohamed, A., & Ahmad Tarmizi, S. A. (2021). Washback impact of the MUET: The before and after effect of a high-stake university English test in Malaysia. *International Journal of Learning, Teaching and Educational Research*, 20(8), 1–21.
- Nguyen, H. T. M. (2023). The washback of the International English Language Testing System (IELTS) as an English language proficiency exit test on the learning of final-year English majors. *TESL-EJ*, 27(1), 35–54.
- Nguyen, T. T. M. (2016). Developing global English language proficiency: Challenges and opportunities. *TESOL Quarterly*, 50(3), 673-683.
- Owusu, S. (2021). Washback effect of high-stakes English language tests on the learning behaviors of Ghanaian ESL learners. *Journal of Education and Practice*, 12(30), 78–92.
- Puspitasari, M. (2024). Social and cultural factors beyond washback of the national examination: Exploring the impact of high-stakes testing in Indonesian secondary schools from teachers' experiences and sociocultural factors. *The Qualitative Report*.
- Putra, R., & Faridi, A. (2021). The washback of formative assessment in EFL instructions. *English Education Journal*.
- Sadeghi, K., Ballıdağ, A., & Mede, E. (2021). The washback effect of TOEFL iBT and a local English proficiency exam on students' motivation, autonomy and language learning strategies. *Heliyon*, 7(10), e08135.
- Sağlam, A. L. G., & Farhady, H. (2019). Can exams change how and what learners learn? Investigating the washback effect of a university English language proficiency test in the Turkish context. *Advances in Language and Literary Studies*, 10(1), 177–188.
- Smith, J., & Johnson, R. (2018). Language proficiency and its impact on academic and career success. *Journal of Applied Linguistics*, 45(2), 123-140.
- Taylor, L. (2005). Washback and impact. *English Language Teaching Journal*, 59, 154-155.
- Van Tubergen, F., & Kalmijn, M. (2022). Language proficiency and educational

- success: Evidence from immigrant children. *Journal of Multilingual and Multicultural Development*.
- Wang, C. (2023). Washback effects of the new CET4 listening test on college English listening teaching and learning. *Advances in Social Sciences Research Journal*.
- Watanabe, Y. (2013). The National Center Test for University Admissions. *Language Testing*, 30(4), 565-573.
- Xie, Q., & Andrews, S. (2022). The impact of rubric transparency on university-level EFL assessments: A study on targeted preparation and student performance. *Language Testing*, 39(3), 345–367.
- Yue, Y. F. (2022). Global Competency Cultivation by Converting College English into Information Literacy Curriculum. In *Inted2022 Proceedings* (pp. 1791-1798). IATED.