

FROM TESTS TO TEACHING: EXPLORING THE WASHBACK EFFECT IN LANGUAGE PEDAGOGY

WMPYB Rathnayake¹

Abstract

The washback effect and the influence of assessments on teaching and learning play a crucial role in shaping instructional practices and nurturing learner experiences. This study explores the evolution of washback over the past four decades, revealing its transition from negative connotations associated with high-stakes testing to a more nuanced understanding of positive and negative pedagogical effects. Narrated along the evolution, conceptualisations, and future directions of the washback effect as applicable in language pedagogy, it emphasises the impact of washback on educational practices and outcomes, enlightening the potential space available for re-engineering language assessment practices. A traditional narrative review enriched by a Global South critical approach, purposely utilised to harness the limits inherent to Global South academia, serves as the methodology. The study interprets existing literature, recognises the existing gaps, and provides a comprehensive overview by ascertaining various perspectives while integrating recent advancements in the field. Key findings confirm that early conceptualisations of washback were primarily concerned with its adverse effects, which have eventually diverged to highlight its potential to foster positive educational practices. Emphasising the importance of aligning assessments with curriculum goals while/and improving teachers' assessment literacy, the study further reveals the modern tendencies of washback literature. Offering directives for future research, the study emphasises the need for greater focus on understanding how washback influences educational outcomes through cognitive processes, equity issues, and stakeholder experiences. Examining the interplay between assessment design and teaching practices will be crucial in optimising the washback effect to enhance learning. The study recommends that educators and policymakers design assessments that measure learning and support and improve the educational process. By addressing the evolving nature of washback through the integration of contemporary insights, the study aims to contribute to more effective, equitable and humanistic assessment practices in language education.

Keywords: Assessment as Learning (AaL), Assessment for Learning (AfL), Global South, Language Pedagogy, Washback Effect

¹ Senior Lecturer, Department of Languages, University of Moratuwa, Sri Lanka

Email: prasannar@uom.lk



<https://orcid.org/0000-0003-2835-3249>



[The Journal of Desk Research Review and Analysis](#) © 2024 by [The Library, University of Kelaniya, Sri Lanka](#) is licensed under [CC BY-SA 4.0](#)

Introduction

Evaluations, assessments, or tests are indicators of learners' development and remain essential components in pedagogy. They are pivotal in determining the trajectory of a learner's educational journey. The impact of these assessments extends beyond mere measurement; they influence teaching methods and learning outcomes—a phenomenon known as the washback effect. The washback effect, with no exception in the context of language pedagogy, plays a decisive role in shaping instructional practices and sustaining the learning experience.

When an assessment is administered toward the end of the educational journey, learners receive feedback only after completing the entire course, which impedes their ability to relate the feedback to their learning trajectory. This significantly impacts the effectiveness of their learning process; therefore, timely insights are crucial for the incremental development of learners. Frequent assessments throughout the learning process provide timely feedback, enabling learners to navigate and invest more effectively in their educational endeavours. These two dimensions of assessment, often taxonomised as assessment of learning (AoL) or summative assessments and assessment for learning (AfL) or formative assessments, are becoming foundational concepts in assessment literature.

Assessment for Learning (AfL) originated from the reconceptualisation of classroom-based formative assessment (Black & Wiliam, 1998; Rea-Dickins, 2001), which was defined by the Assessment Reform Group (2002) as a process for gathering and interpreting evidence to guide learners and teachers in understanding learning progress, setting goals, and determining strategies to achieve those goals, creates opportunities for collaborative feedback and teacher-student interactions that promote student reflection, awareness, realistic goal setting, and autonomy (Leung & Mohan, 2004). AfL practices are gaining significance as supporting student motivation and providing teachers with valuable insights into student thinking and understanding (Cheng & Fox, 2017; Wiliam, 2011), thus gaining a universal consensus on the critical role of AfL in enhancing learning (Cizek, 2010; Vogt & Tsagari, 2014). Within the AfL subset, Assessment as Learning (AaL) emphasises students as active agents in the assessment process, allowing them to connect assessment with their prior learning and set individualised progress goals (Earl, 2013). Self-assessment inventories, portfolio collections, and peer assessments, which enable students to use assessment to further their learning (Cheng & Fox, 2017), remain such examples of AaL, which is closely related to self-regulated learning (SRL), where students actively connect new information with prior understandings and adapt their thoughts and actions to achieve personal goals (Zimmerman, 2000; Andrade & Brookhart, 2019).

Regardless of the terminology used, understanding the comprehensive effects of assessments on educational contexts is essential. This comprehension is best achieved through the lens of the washback effect, which underscores the interplay between assessments, teaching, and learning. Alderson and Wall (1993) propose that the distinction between significant and insignificant tests can be delineated through the criterion of washback. They define washback as the effect of tests on various aspects of the educational process, encompassing teachers, students, content, methods, pacing, sequencing, depth, degree, and attitudes towards teaching and learning. Washback is widely acknowledged as the influence of tests on teaching and learning (Pearson, 1988; Hughes, 1989; Bailey, 1996; Alderson & Wall, 1993).

Bailey (1996) acknowledges that washback "can be either positive or negative, promoting or hindering the achievement of educational goals" (p. 269). Depending on its orientation, a test can significantly impact daily teaching and learning practices, thus guiding educators in enhancing the process. Hughes (1989) argues that positive washback can be enhanced by designing direct, cost-effective, objective-driven, and criterion-referenced tests. Negative washback, on the other hand, arises from teaching

focused solely on test preparation, neglecting the development of essential skills expected of learners, or from misalignment between curriculum goals and examination content (Cheng & Curtis, 2004).

Cheng and Curtis (2012) distinguish between washback and impact, with washback having a micro-focus and narrow scope, examining effects within the classroom. In contrast, the impact has broader implications beyond the classroom. It analyses assessment effects on educational systems and society, involving various stakeholders.

To understand how assessments can drive educational efficacy and inform pedagogical strategies, this review examines the relationships within the web of test and assessment practices in language pedagogy. The significance of this interplay has been the focus of many studies, highlighting the need for educators and policymakers to consider the broader implications of testing within educational systems. In this context, understanding the washback effect is essential for optimising language education.

This study, amid the inherent sociopolitical and economic consequences affecting global southerners' deprived access to academic resources, endeavours to fill the void of context-specific reviews on washback as a critical concept of educational significance in language pedagogy. It questions: *How has washback evolved over the past four decades? What conceptualisations have nurtured its trajectory? Moreover, What epistemological directives guide its present and future directions?* By narrating the saga ***From Tests to Teaching: Washback Effect in Language Pedagogy***, this study offers directives for language pedagogy practitioners to better design their educational artefacts. Attempted via a Global South Critical Approach to the literature review, this study of the washback literature explores the historical evolution of the term 'washback' from its original interpretations to the modern and sophisticated connotations attributed to it, thereby defining the concept logically for its constructive application in language pedagogy. The review delves into the trajectory of washback research and its utility for pedagogical innovation, with a keen focus on language pedagogy.

Methodology

This review employs the traditional narrative review method, which qualitatively interprets existing literature without aiming for generalisation (Sylvester et al., 2013). This method involves summarising various viewpoints on a topic, occasionally overlooking certain studies to emphasise specific perspectives (Davies, 2000; Green et al., 2006). It is particularly valued in education for its up-to-date nature and access to peer-reviewed sources, as it helps synthesise research ideas by identifying gaps (Green et al., 2006). Potential subjectivity and lack of clear criteria are drawbacks of this method. However, its strength lies in gathering a range of literature to provide context for new research (Cronin et al., 2008). Recent developments, such as the structured frameworks introduced by Levy and Ellis (2006), which enhance systematic literature search, data extraction, and analysis, have been incorporated into this method. This adaptation leverages the traditional narrative review's strengths in guiding educational practices and research agendas while emphasising methodological rigour (Paré & Kitsiou, 2017). Given these merits, this study employs a traditional narrative review methodology driven by a Global South critical approach.

This study uniquely employs a Global South critical approach-driven traditional narrative review methodology. This adaptation addresses the inherent deficits in access to leading academic publications in Global South localities. Originating from Colombo, a city in the Global South (Sivasundaram, 2023), this review highlights significant inequities faced by researchers in these regions. Barriers such as limited resources, insufficient support for specific subject areas, high subscription costs, and

complexities in managing access hinder the Global South academia's accessibility to scholarly resources (Gwynn, 2019).

Despite the rise of Open Access, about two-thirds of scholarly literature remains behind paywalls, disproportionately affecting researchers in Low- and Lower-Middle-Income Countries (Piwowar, 2018). Therefore, this review aims to overcome these barriers as a Global South initiative, employing an adapted authentic review sources access strategy.

Following Saunders and Rojon's (2011) checklist for critical literature reviews, this review ensures that relevant and significant research is identified and included, discussed logically, references recognised experts, contextualises research aims and objectives, includes supporting and opposing research with clear reasoning, distinguishes fact from opinion, incorporates recent research, and references all sources thoroughly.

The methodological rigour of this review categorises it as a Global South Critical Approach to a traditional narrative review. The process includes accessing credible reference guides (Mehebbi & Coombe, 2021), recent encyclopedia entries on washback (Chapelle, 2020), and conducting back-and-forth cross-referencing of seminal citations. Further cross-referencing with recent handbooks (Fulcher & Harding, 2022) and a Google Scholar search on "washback" research for works published after 2021 was conducted to select academic sources for the study. Using this corpus of authentic literature, the final synthesis was fine-tuned according to Saunders and Rojon's (2011) critical review checklist.

Results and Discussion

Narrated using a Global South critical approach-driven traditional narrative review method, the present study, in the order of the research questions that set its base, includes the saga "From Tests to Teaching: Washback Effect in Language Pedagogy." In this narrative, the higher-order assertions arrived upon the discussion will be summarised and presented following the discussion of each research question. At the end of this section, I highlight a critical gap in the assessment literature, particularly in washback studies related to the Sri Lankan context, despite this not being the primary focus of the current narrative.

R.Q.1: How has washback evolved over the past four decades?

Early Developmental Milestones

The washback concept has undergone significant evolution, reflecting changing perspectives on the impact of assessments on education. Initially, washback was linked to various terms such as "test impact" (Bachman & Palmer, 1996), "consequential validity" (Messick, 1989), "measurement-driven instruction" (Popham, 1987), "curriculum alignment" (Shepard, 1990), and "backwash" (Biggs, 1995). Despite the variety of terminology, "washback" emerged as the dominant term in academic discourse, particularly within applied linguistics (Cheng et al., 2004; Green, 2007; Hawkey, 2006).

Historically, the idea of washback dates back to the "encroaching power" of high-stakes examinations in the late 19th century (Latham, 1877). Initially, washback was viewed negatively as a detrimental effect of high-stakes tests. However, the concept broadened over time to encompass positive and negative influences. Early recognition of its potential positive effects emerged in the mid-20th century (Wilkinson, 1965), highlighting how strategically designed assessments could enhance desirable educational practices. The scope of washback has expanded to include individual, classroom, school, system, and societal impacts (Wall, 1997).

Advancements in Washback Research

Systematic research into washback began with Alderson and Wall (1993), who proposed 15 hypotheses on the potential impacts of tests. This seminal work paved the way for subsequent frameworks. Critical components of washback include participants, processes, and products, as Hughes (1993) identified. Bailey (1996) introduced a dynamic model of washback that depicts interactions within educational settings. Applying this model to practice, Cheng's studies in Hong Kong (1997, 1998, 1999) revealed gaps in stakeholder understanding and highlighted the need for improved teacher assessment literacy. Green (2007) further advanced the field with a learner-centred model that links washback to test validity.

Examining significant tests such as TOEFL (Alderson & Hamp-Lyons, 1996), IELTS (Green, 2007), and university entrance exams in Japan (Watanabe, 1996) and China (Qi, 2005), washback research has expanded to global contexts. These studies highlighted the diverse impacts on teaching practices and student outcomes, emphasising the need to align testing methods and educational goals. Recent research, including studies by Rahman et al. (2021) and Athiworakun and Adunyarittigun (2022), underscores the complexity of washback, showing that it involves not only aligning teaching with test objectives but also integrating curriculum goals and formative assessment practices. This evolution reflects a growing recognition of the interconnectedness between assessment, teaching, and learning and the importance of teachers' understanding of curriculum and assessment literacy.

Evolution and Dimensions of Washback Research

The initial negative perception of washback has evolved to encompass both positive and negative effects. Hughes (1989) advocated for assessments promoting desirable skills, while Xin (2021) emphasised optimising tests for teachers and learners. Building on this, researchers have identified six key dimensions of washback: direction (positive or negative), extent (reach across different stakeholders), intensity (variation with test stakes), intentionality (planned or unintended effects), length (short-term or enduring impacts), and specificity (precise domains affected) (Bachman & Palmer, 2010; Green, 2007; Watanabe, 2004; Cheng, 2005).

Marking a shift from a cause-and-effect perspective to more nuanced approaches, the understanding of washback has evolved to acknowledge its complexity, as observed by Cheng and Sultana (2022). They highlighted the importance of context-specific investigations and robust frameworks to grasp the multi-layered nature of washback studies. Research in diverse contexts, such as Sri Lanka (Umashankar, 2017) and the ACTFL Assessment (Vyn, 2019), has replicated existing frameworks, enhancing the understanding of washback's global impact. Studies by Wong Katagiri (2023) and Samaie and Valizadeh (2023) reveal the influence of sociocultural factors and the critical role of teacher agency in shaping washback, moving the discourse beyond simplistic cause-and-effect models to explore the broader educational landscape.

Towards the Future

Cheng and Sultana (2022) recommend that future washback research focus on stakeholders' cognition to understand how washback is experienced and its impact on educational outcomes. Looking ahead, it remains crucial to address the connection between teachers' assessment literacy and washback, explore cognitive processes in student learning, and consider issues of equity and fairness in testing. Innovative methods are deemed necessary to capture stakeholders' cognitive experiences and reveal more profound insights into the impact of washback on teaching and learning.

Summary of the Evolution of Washback Over the Past Four Decades

Over the past four decades, washback research has transitioned from negative perceptions to a comprehensive understanding of its dual nature. Beginning with Alderson and Wall's foundational work in 1993, the field has evolved through significant theoretical advancements and marked versatile global expansions. Theoretical models have deepened the understanding of washback's interaction with validity and educational outcomes. Recent research emphasises methodological advancements and socio-political dimensions, reflecting washback's role in shaping global educational practices. Focusing on stakeholders' cognitive experiences and addressing gaps through innovative research methods will be key to enhancing the quality and fairness of educational assessments.

R.Q.2: What conceptualisations have nurtured the washback research trajectory?

Although not significantly disconnected from the trajectory of washback as a concept in pedagogy, which has been the central concern of this narrative thus far, this section will trace the conceptual frameworks that have guided and shaped the evolution of washback literature. Specifically, I will explore the critical milestones in the literature that have evolved from foundational concepts of washback, examining how these concepts have been applied and expanded upon in research. This includes the theoretical developments and conceptual models, as well as the recent global perspectives and contemporary challenges that have emerged and those yet to be encountered.

Foundational Concepts and Early Frameworks

The cornerstone query, "Does washback exist?"—pioneeringly investigated by Alderson and Wall (1993)—concluded that "yes, washback exists, but it is complicated" (Fulcher & Harding, 2022, p. 5). They introduced initial hypotheses on the impacts of tests, laying the groundwork for systematic washback research in language testing. Their work theorised 15 hypotheses regarding washback, highlighting the influence of tests on teachers, students, content, method, rate, sequence, degree, depth, and attitudes toward teaching and learning.

Hughes (1993) identified washback's key components (producers): participants, processes, and products. Bailey (1996) later expanded these into a comprehensive model illustrating dynamic interactions within educational settings. Bailey's model, built upon Hughes' three key producers of washback and Alderson and Wall's 15 hypotheses, demonstrates not only that tests stimulate products through the participants and the processes of teaching and learning but also that feedback provided by participants may, in turn, bring changes to the test. This model visualised the dynamic interactions and feedback loops within educational settings, giving scholars room for exploration. Such developments in the 1990s conceptualising washback laid the foundation for future nurture that was later received.

Application and Expansion

Bailey's model gained wider acceptance when Cheng (1997, 1998, 1999) applied it in his subsequent studies to specific assessments, such as the Hong Kong Certificate of Education Examination. These explorations revealed gaps in teacher assessment literacy and the need for alignment with educational objectives. Furthermore, studies on major tests such as TOEFL (Alderson & Hamp-Lyons, 1996), IELTS (Green, 2007), and regional exams in Japan and China (Watanabe, 1996; Qi, 2005) demonstrated diverse effects on teaching practices and student outcomes, confirming the conceptual richness of the washback model in practice. Cheng and Sultana acknowledge that Bailey's model "visualised the non-linear relationship among the education components and illustrated the complexities that washback research carries forward." However, they elaborate that Bailey did not focus on the "processes" between products and participants (Cheng & Sultana, 2022, p. 137).

Theoretical Developments and Conceptual Models

Green (2007) introduced a learner-centred model of washback, focusing on direction, variability, and intensity while linking it to test validity. Green's (2007) model, which includes the three dimensions of direction, variability, and intensity, aimed to address the gap in understanding students' learning. It "offered a comprehensive model focusing on learners and learning linking back to test validity" and thus remains "the most dominant learner-centred washback model" (Cheng & Sultana, 2022, p.137).

Wall and Horák (2007) identified two types of washback studies: those examining the influence of tests already in operation and those exploring the effects of newly introduced tests. Wall (2012) prescribes two distinct approaches for each type. The emergence of theoretical models has integrated insights from Wall and Horák's longitudinal studies and assessment developers' explicit declarations of intended washback effects (Green, 2020).

Global Perspectives and Contemporary Challenges

Evolving through such milestones, as noted by Cheng and Sultana (2022), a notable shift is evident in washback research towards cognitive-oriented approaches. Concurrently, the integration of validity frameworks to understand the complex, non-linear relationship between washback and test validity remains significant. The importance of innovative research methods to capture stakeholders' cognitive experiences and thereby gain deeper insights into the impact of washback on teaching and learning represents an emerging epistemological insight. Recent studies, such as those by Rahman et al. (2021), have expanded this view by emphasising teachers' need for a deep understanding of curriculum objectives and robust assessment literacy. Athiworakun and Adunyarittigun (2022) highlight the motivational role of high-stakes tests in driving pedagogical innovation. Wong Katagiri (2023) introduces the idea of mediating factors, such as peer influence and sociocultural context, which significantly shape students' learning choices. Additionally, Samaie and Valizadeh (2023) explore the concept of assessment as learning (AaL), underscoring the integration of formative assessment into teaching practices to promote a more supportive and adaptive learning environment. These conceptualisations shift the landscape towards viewing washback as a complex, dynamic phenomenon influenced by various educational, social, and cultural factors.

Summary of Key Conceptualizations Shaping the Evolution of Washback Research

Our trace of the evolution of washback research reveals a progression from foundational studies to complex, nuanced explorations. Initial hypotheses by Alderson and Wall (1993) were expanded by Hughes (1993) and Bailey (1996) into models illustrating dynamic educational interactions. Cheng's application of Bailey's model highlighted gaps in teacher assessment literacy and the need for alignment with educational objectives.

Theoretical advancements like Green's learner-centred model linked washback to test validity, addressing variability and intensity. Wall and Horák's distinctions between studies on existing and new tests refined research methodologies. The shifts toward cognitive-oriented approaches and integrating validity frameworks reflect the maturation of washback research. Recent studies underscore a holistic understanding of educational dynamics, highlighting the necessity for robust teacher assessment literacy, the motivational role of high-stakes tests, the influence of sociocultural contexts, and the integration of formative assessment, reflecting washback as a complex, multi-faceted phenomenon. These developments underscore the importance of innovative methods to capture stakeholders' cognitive experiences, enhancing our understanding of the complex interplay between washback, teaching, and learning.

R.Q.3: What epistemological directives guide present and future directions of washback literature?

Washback, or the influence of testing on teaching and learning, is a multi-faceted construct that has garnered significant scholarly attention. This inquiry into the epistemological directives identifiable in washback research involves examining the nature, scope, and limitations of knowledge and the methods through which knowledge is acquired. This entails reviewing the existing literature through the lenses of the nature of knowledge, sources of knowledge, justification of knowledge, and methodologies for accessing knowledge. By doing so, I critically reflect on the processes and assumptions underlying the scientific inquiry into washback and synthesise key findings from the literature to elucidate the epistemological directives guiding washback researching the present and future research.

The necessity to comprehensively evaluate washback by addressing its various dimensions, which collectively define its complexity and scope (Bailey, 1996; Cheng, 1997; Watanabe, 1996), was a concern of Xu and Liu (2018). Shedding light on this directive, researchers have attempted to perfect the epistemological framework for washback studies over the past four decades. Synthesised below is how these dimensions of washback have portrayed washback as a robust construct.

Direction of Washback: Positive vs. Negative

The literature identifies two primary directions of washback: positive and negative. Positive washback occurs when tests are well-designed, facilitating and motivating effective teaching and learning practices. Equally, negative washback arises from poorly designed tests that may lead to undesirable educational practices (Bachman & Palmer, 2010; Green, 2007). Based on the evidence generated in washback studies, such is the basis of any test that may turn out to be a well-designed or poorly-designed test. This contrast serves as a direction for researchers to scrutinise test design critically, aiming to enhance positive outcomes while mitigating negative ones.

Scope and Extent of Washback

The scope of washback is a seminal concern among its dimensions, as it connects various stakeholders of any test. Whether functioning as a gatekeeper or an enabler, a test substantially affects its stakeholders, influencing both the process and the product of education. The literature explores the extent of washback and its reach across different educational stakeholders, including students, teachers, and administrators (Watanabe, 2004). High-stakes tests are often highlighted in the literature for their extensive washback, impacting broader educational contexts. This aspect of research looks at how washback appears at different levels of the educational system, ranging from individual classrooms to entire institutions. It examines how washback operates at universal, regional, national, institutional, disciplinary, or module levels. Understanding the scope and extent of washback is crucial for properly contextualising tests. This approach, known as glocalisation, ensures that tests are relevant and suitable for the specific contexts in which they are used.

Intensity of Washback

The intensity or strength of washback effects is a crucial aspect to explore. High-stakes tests, such as summative assessments, often exert more pronounced impacts on teaching and learning practices (Cheng, 2005; Green, 2007). In contexts where summative assessments, like final exams, are heavily emphasised and play a significant role, their impact on teaching and learning can be markedly different from contexts where formative assessments are prioritised. Summative assessments generally have high stakes and can significantly influence educational practices, while formative assessments focus on providing ongoing feedback to enhance learning processes.

The varying emphasis on Assessment of Learning (AoL) and Assessment for Learning (AfL) can lead to different intensities of washback, thus affecting the utility of these assessment formats and their potential to support the teaching-learning process. Understanding these differences is essential for analysing how various types of assessments impact educational outcomes. Examining how different levels of stakes influence educational behaviours and attitudes can help future test designers balance these effects more effectively.

Intentionality:

Intended vs. Unintended Washback

Washback research identifies two main types of outcomes: intended and unintended consequences. Intended outcomes help guide test development towards positive results, while unintended outcomes can also significantly impact, sometimes more than expected. Washback can be intended, meaning it aligns with the goals set by test designers, or unintended, which includes unexpected effects like increased student anxiety or excessive test preparation (Cheng, 2005).

This aspect of research supports a comprehensive approach to evaluating the impact of tests, encouraging researchers to anticipate and recognise both expected and unexpected outcomes.

Temporal Length of Washback

The duration or how long the effects of washback last is an essential area of study. The length of washback effects can vary; some impacts may be short-lived, while others can last longer (Bachman & Palmer, 2010). For example, motivational effects might fade quickly, but habits formed during test preparation can continue. Longitudinal studies can track these changes over time, providing insights into both immediate and long-term educational outcomes.

Specificity of Washback

Among the various dimensions of washback, specificity is significant because many educational interventions are designed with precise objectives in mind. This dimension focuses on the areas or aspects affected by washback. It warrants further investigation across different educational contexts and scales. Researchers can enhance our understanding of how washback influences various levels of pedagogy by exploring specificity in greater detail, from individual classrooms to broader educational systems. Washback effects can be general, influencing broad educational processes, or specific, linked to particular tests or test types (Cheng, 2005). Investigating both general and specific effects facilitates a comprehensive understanding of how different tests uniquely impact teaching and learning (Watanabe, 1996).

Towards a Comprehensive Understanding of Washback

The epistemological directives ascertained from these dimensions emphasise a multi-dimensional and nuanced approach to washback research. Having cleared the roots of such epistemology of washback studies, what is presented below includes the directions of washback research. Future studies should continue critically analysing test design, enhancing positive washback and reducing negative impacts, thereby contributing to more effective and equitable educational assessments. Additionally, examining washback across broad and specific contexts, including different educational levels and environments, will provide a holistic view of its influence, ensuring a comprehensive understanding of test impacts. Longitudinal studies are necessary to grasp the enduring effects of testing, which will inform better practices for sustained positive educational outcomes. Anticipating and mitigating unintended negative consequences through a proactive approach will lead to more responsible test administration and use.

Integration of Contemporary Studies

Recent studies, such as those by Maksud Ali and Hamid (2023), go beyond just focusing on how tests drive teaching methods. They explore how teachers and students deal with and manage the pressures of assessments. This view emphasises the importance of understanding how teachers and students interact and adapt to assessments. Samaie and Valizadeh (2023) stress the value of ongoing formative assessments that help develop self-regulated learning and metacognitive skills. Additionally, Wong Katagiri (2023) shows how sociocultural factors influence learning, highlighting the need to consider the relationship between assessments, learning, and social contexts.

Together, these insights support a more complete and context-sensitive approach to washback research. This helps improve teaching methods and policies, leading to better educational outcomes. By including these perspectives, washback research becomes richer and provides a deeper understanding of how testing affects education. This approach adds to academic discussions and helps improve educational systems worldwide.

Summary of Guiding Epistemological Principles for Washback Literature

Guided by epistemological directives that scrutinise knowledge's nature, scope, and limitations, Washback has become a significant focus of scholarly inquiry. Researchers critically reflect on the processes and assumptions underlying washback by examining the literature through the lenses of the nature of knowledge, sources of knowledge, justification of knowledge, and methodologies for accessing knowledge. The dimensions of washback encompass its direction (positive vs negative), scope and extent, intensity, intentionality (intended vs unintended), temporal duration, and specificity. Future research should prioritise the critical analysis of test design, investigate washback across diverse educational contexts, explore longitudinal effects, and anticipate unintended consequences to promote more effective and equitable educational assessments. Aligning with a holistic approach to washback research, studies emphasise the importance of teacher and student agency, advocate for formative assessment integration, and highlight sociocultural factors. Adopting a multi-dimensional approach will enhance the understanding and management of washback in educational settings, ultimately fostering improved pedagogical practices and outcomes.

Globalizing Washback Studies to Optimise the Sri Lankan Learning Experience

In the context of Sri Lankan language education, the washback effect may be influenced by the widespread issue where students fail to learn the skills outlined in the curriculum, as observed in many educational systems globally (Aturupane et al., 2022; Glewwe & Kremer, 2006; Glewwe & Muralidharan, 2016; Damon et al., 2019). In examining the washback effects in language education, Sri Lanka emerges as a significant context. As highlighted by Wang (2021), such notable studies include Wall and Alderson's (1993) investigation of O-level English as an International Language through various qualitative methods and Iyer's (2015) exploration of English language tests at Sri Lankan universities. Additionally, our review captured the essence of Umashankar's (2017) study on the washback effects of speaking assessment on English teaching in Sri Lankan schools.

A critical gap exists in the literature regarding the mismatch between testing and curriculum objectives in Sri Lanka, with a notable lack of authentic studies on the washback effect; however, this narrative offers insights for future research by highlighting the potential for re-engineering washback effects in tertiary-level language pedagogy to address broader educational challenges.

Conclusions and Recommendations

This study employs a traditional narrative review method to investigate the evolution, conceptualisations, and epistemological foundations guiding washback research in language pedagogy.

It explores significant aspects of washback over the past four decades, offering insights into its historical development, theoretical frameworks, and global perspectives. The review leads to several conclusions and recommendations, which are presented below.

Conclusions

Washback has evolved along a constructive trajectory from being perceived as a negative consequence of assessments to a nuanced understanding encompassing both positive and negative impacts. This saga, from the foundational studies by Alderson and Wall (1993) to contemporary global perspectives, reflects a comprehensive exploration of its complexities and dynamics.

The conceptual frameworks of washback have been diverse and robust. They range from Alderson and Wall's early hypotheses to Bailey's dynamic model and Green's learner-centered approach. These frameworks have expanded our understanding of how assessments influence educational practices across various levels.

The study identifies key epistemological directives shaping washback research, including the direction (positive vs. adverse effects), scope and extent, intensity, intentionality, temporal duration, and specificity of washback effects. These directives underscore washback's complex and multi-dimensional nature as a research construct.

There is a critical gap in the literature concerning the mismatch between testing and curriculum objectives in Sri Lanka, a widespread issue in education. However, a noticeable lack of authentic and critical studies focused on the washback effect could be pivotal in addressing long-standing challenges in general education. Initiatives to explore this area might be more feasible at the tertiary level. While our review here slightly deviates from this focus, it still offers insights relevant to the Sri Lankan context.

Recommendations

Directives for future research highlight the need to prioritise methodological innovation, emphasising the need for innovative research methods capable of capturing the cognitive experiences of stakeholders affected by washback. Conducting longitudinal studies to track the enduring effects of washback on educational outcomes and practices is essential for advancing the field.

Significant potential exists for contextually specific investigations into washback effects across diverse educational settings.

Regarding policy and practice implications highlighted in the review, there is a pressing need to enhance teacher assessment literacy to optimise the positive impacts of washback and mitigate negative consequences.

Future trends in washback research should integrate sociocultural perspectives to understand better how cultural contexts influence assessment practices and outcomes.

Promoting the integration of formative and summative assessment practices is crucial at policy and practice levels to effectively support student learning and development.

Limitations of the Study

It is essential to critically examine the washback effects, particularly in the context of Sri Lankan public examinations and assessments, to gain valuable insights. This line of research is crucial and should be pursued further. Due to the limited literature on the Sri Lankan narrative of washback in language pedagogy and my focus on exploring washback within a broader theoretical and conceptual framework, I have deliberately narrowed my review to align with the study's objectives. This decision has created a limitation, highlighting the need for future research to explore the washback effect's critical and timely significance in aligning Sri Lanka's general education system outcomes. I leave this as a gap for future scholars to address.

Acknowledgement

I am grateful for the invaluable support of ChatGPT and its insightful proofreading and fact-checking assistance. I also appreciate the University of Moratuwa Library staff for their exceptional help in providing access to essential academic resources.

References

- Alderson, J., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14(2), 115–129.
- Alderson, J., & Hamp-Lyons, L. (1996). TOEFL preparation courses: A study of washback. *Language Testing*, 13(3), 280–297.
- Andrade, H. L., & Brookhart, S. M. (2019). Classroom assessment as the co-regulation of Learning. *Assessment in Education: Principles, Policy & Practice*, 27(4), 350–372. <https://doi.org/10.1080/0969594X.2019.1571992>
- Assessment Reform Group. (2002). *Assessment for learning: 10 principles*. Assessment Reform Group. Retrieved from www.aai.org.uk/storage/medialibrary/o_1d8j89n3u1n0u17u91fdd1m4418fh8.pdf
- Athiworakun, C., & Adunyarittigun, D. (2022). Investigating washback effects on teaching: A Case study of an exit examination at the higher education level. *LEARN Journal*, 15(2), 776–801.
- Aturupane, H., Glewwe, P., Utsumi, T., Wisniewski, S., & Shoji, M. (2022). The impact of Sri Lanka's school-based management programme on teachers' pedagogical practices and student learning: Evidence from a randomised controlled trial. *Journal of Development Effectiveness*, 14(4), 285–305. <https://doi.org/10.1080/19439342.2022.2029540>
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and Developing helpful language tests*. Oxford University Press.
- Bachman, L., & Palmer, A. (2010). *Language assessment in practice*. Oxford University Press.
- Bailey, K. (1996). Working for washback: A review of the washback concept in language testing. *Language Testing*, 13(3), 257–279.
- Biggs, J. (1995). Assessing for learning: Some dimensions underlying new approaches to educational assessment. *Alberta Journal of Educational Research*, 41(1), 1–17.
- Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom Assessment. *Phi Delta Kappan*, 92, 81–90. <https://doi.org/10.1177/003172171009200119>
- Chapelle, C. (2020). *Argument-based validation in testing and assessment*. Sage.
- Cheng, L. (1997). How does washback influence teaching? Implications for Hong Kong. *Language and Education*, 11, 38–54. <https://doi.org/10.1080/09500789708666717>
- Cheng, L. (1998). Impact of a public English examination change on students' perceptions and attitudes toward their English learning. *Studies in Educational Evaluation*, pp. 24, 279–301. [https://doi.org/10.1016/S0191-491X\(98\)00018-2](https://doi.org/10.1016/S0191-491X(98)00018-2)
- Cheng, L. (1999). Changing assessment: Washback on teacher perceptions and actions. *Teaching and Teacher Education*, 15, 253–271. [https://doi.org/10.1016/S0742-051X\(98\)00046-8](https://doi.org/10.1016/S0742-051X(98)00046-8)
- Cheng, L. (2005). *Changing language teaching through language testing: A washback study*. Cambridge University Press.
- Cheng, L. (2008). Washback, impact and consequences. In E. Shohamy & N. H. Hornberger

- (Eds.), *Encyclopedia of language and education* (pp. 2479–2494). Springer.
- Cheng, L. (2014). Consequences, impact, and washback. In A. J. Kunnan (Ed.), *The Companion to language assessment* (pp. 1–14). John Wiley & Sons.
- Cheng, L. (2018). Geopolitics of assessment. In S. Abrar-ul-Hassan (Ed.), *TESOL Encyclopedia of English language teaching: Teaching English as an international language*. John Wiley & Sons.
- Cheng, L., & Curtis, A. (2012). Test impact and washback: Implications for teaching and Learning. In G. Fulcher & F. Davidson (Eds.), *The Routledge Handbook of language testing* (pp. 89–95). Routledge.
- Cheng, L., & Fox, J. (2017). *Assessment in the language classroom*. Palgrave Macmillan.
- Cheng, L., & Sultana, N. (2022). Washback: Looking backwards and forward. In G. Fulcher & L. Harding (Eds.), *The Routledge Handbook of Language Testing* (2nd ed., pp. 136–152). Routledge.
- Cheng, L., Watanabe, Y., & Curtis, A. (Eds.). (2004). *Washback in language testing: Research Contexts and methods*. Lawrence Erlbaum Associates.
- Cizek, G. J. (2010). An introduction to formative assessment: History, characteristics and challenges. In H. L. Andrade & G. J. Cizek (Eds.), *Handbook of formative assessment* (pp. 3–17). Routledge.
- Cronin, P., Ryan, F., & Coughlan, M. (2008). Undertaking a literature review: A step-by-step Approach. *British Journal of Nursing*, 17(1), 38–43.
<https://doi.org/10.12968/bjon.2008.17.1.28059>
- Damon, A., Glewwe, P., Wisniewski, S., & Sun, B. (2019). What education policies and Programmes affect learning and time in school in developing countries. A review of evaluations from 1990 to 2014. *Review of Education*, 7(2), 295–387.
<https://doi.org/10.1002/rev3.3123>
- Davies, P. (2000). The relevance of systematic reviews to educational policy and practice. *Oxford Review of Education*, 26(3–4), 365–378. <https://doi.org/10.1080/713688543>
- Earl, L. M. (2013). *Assessment as learning: Using classroom assessment to maximise student learning*. Corwin Press.
- Fulcher, G., & Harding, L. (2022). *The Routledge Handbook of language testing* (2nd ed.). Routledge.
- Glewwe, P., & Kremer, M. (2006). Schools, teachers, and education outcomes in developing Countries. In E. Hanushek & F. Welch (Eds.), *Handbook of the economics of education* (Vol. 2, pp. 945–1017). North Holland: Elsevier.
- Glewwe, P., & Muralidharan, K. (2016). Improving education outcomes in developing Countries: Evidence, knowledge gaps and policy implications. In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbook of the economics of education* (Vol. 5, pp. 653–743). North Holland: Elsevier.
- Green, A. (2007). *IELTS washback in context: Preparation for academic writing in higher education*. Cambridge University Press.
- Green, A. (2020). Washback in language assessment. In C. Chapelle (Ed.), *The concise Encyclopedia of applied linguistics* (pp. 1159–1193). Wiley-Blackwell.
- Gwynn, S. (2019). Access to research in the Global South: Reviewing the evidence. Retrieved June 5, 2024, from <https://www.inasp.info/sites/default/files/2019-01/S%20Gwynn%20access%20paper.pdf>
- Hawkey, R. (2006). *Impact theory and practice: Studies of the IELTS test and Progetto Lingue 2000*. Cambridge University Press.
- Hughes, A. (1989). *Testing for language teachers*. Cambridge University Press.
- Hughes, A. (2003). *Testing for language teachers*. Cambridge University Press.
- Iyer, M. S. (2015). Impact of washback in English as second language classrooms – An investigation in the University of Jaffna, Sri Lanka. *ELT Voices-International Journal for Teachers of English*, 5(3), 88–91.
- Latham, H. (1877). *The action of examinations is considered a means of selection*. Deighton, Bell and Company.
- Leung, C., & Mohan, B. (2004). Teacher formative assessment and talk in the classroom

- Contexts: Assessment as discourse and assessment of discourse. *Language Testing*, 21, 335–359. <https://doi.org/10.1191/0265532204lt287oa>
- Levy, Y., & Ellis, T. J. (2006). A systems approach to conduct an effective literature review in Support of information systems research. *Informing Science*, 9, 181–211.
- Mohebbi, H., & Coombe, C. (2021). *Research questions in language education and applied Linguistics: A reference guide*. Springer Nature. <https://doi.org/10.1007/978-3-030-79143-8>
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). American Council on Education; Macmillan.
- Piwowar, H., Priem, J., Larivière, V., Alperin, J. P., Matthias, L., Norlander, B., Farley, A., West, J., & Haustein, S. (2018). The state of open access: A large-scale analysis of the prevalence and impact of open access articles. *PeerJ*, 6, e4375. <https://doi.org/10.7717/peerj.4375>
- Popham, W. J. (1987). The merits of measurement-driven instruction. *Phi Delta Kappan*, 68(9), 679–682. <http://www.jstor.org/stable/20403467>
- Qi, L. (2005). Stakeholders' conflicting aims undermine the washback function of a high--stakes test. *Language Testing*, 22, 142–173. <https://doi.org/10.1191/0265532205lt300oa>
- Rahman, A., Ibna Seraj, P., Kamrul Hasan, M., Namaziandost, E., & Ahmad Tilwani, S. (2021). Washback of assessment on English teaching-learning practice at secondary schools. *Language Testing in Asia*, 11(12), 1-23.
- Rea-Dickins, P. (2001). Mirror, mirror on the Wall: Identifying processes of classroom assessment. *Language Testing*, pp. 18, 429–462. <https://doi.org/10.1177/026553220101800404>
- Richards, J. C., & Renandya, W. A. (2002). *Methodology in language teaching: An anthology of current practice*. Cambridge University Press.
- Shohamy, E. (2001). *The power of tests: A critical perspective on the uses of language tests*. Pearson.
- Stiggins, R. J. (2002). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan*, 83(10), 758–765. <https://doi.org/10.1177/003172170208301005>
- Stiggins, R. J. (2007). *Assessment for learning: An essential foundation for the success of all students*. Assessment Training Institute.
- Stiggins, R. J., Arter, J. A., Chappuis, J., & Chappuis, S. (2006). *Classroom assessment for Student learning: Doing it right—using it well*. Assessment Training Institute.
- Umashankar, S. (2017). *Washback effects of speaking assessment on teaching English in Sri Lankan schools* [Digitised dissertation, University of Bedfordshire].
- Wang, M. (2006). *The impact of high-stakes testing on the quality of English instruction in China*. Routledge.
- Wang, L. (2021). Washback of English language testing on ELL teaching and learning: A literature review. *International Journal of Arts, Humanities & Social Science*, 2(5). <https://doi.org/10.13140/RG.2.2.30530.22726>
- Wang, M. (2008). *The impact of high-stakes testing on educational equity*. Routledge.
- Wang, Y. (2021). Examining the impact of high-stakes testing on teachers' instructional practices: A literature review. *Assessment & Evaluation in Higher Education*, 46(2), 234–248. <https://doi.org/10.1080/02602938.2020.1775139>
- Wiliam, D. (2011). *Embedded formative assessment*. Solution Tree Press.
- Wiliam, D. (2011). *Embedded formative assessment*. Solution Tree Press.