

## WASHBACK EFFECT AND ENGLISH LANGUAGE TESTING IN TÜRKİYE: A CRITICAL REVIEW STUDY

**Suheyyla DEMIRKOL ORAK**

Assist. Prof. Dr. /Fırat University, Faculty of Education, English Language Teaching Department,  
Türkiye /sudemirkol@gmail.com ORCID ID: 0000-0002-0605-6537

### Abstract

Every single teaching and learning action is mentioned with the assessment styles. No matter the style of the examination, it can be either process-oriented or product-oriented, examination has an effect on the test-takers' further learning activities either debilitating or facilitative. Leaning on this immediate fact, the present review study was set to discover primarily the procedural link between the washback effect and English language testing. The secondary aim was to bring a shared understanding of probable effect of the central language examination system in Türkiye on the test-taker by relying on the comparison of Japan and India samples. By doing so, a critical review study was conducted over various search engines ranging from Google Scholar to Web of Science by focusing on the specific keywords. In accordance with the results, it was observed that whilst product-oriented exam styles have debilitating washback effect, process-oriented exam styles have facilitative washback effect on test-takers together with long-lasting learning outcomes.

**Keywords:** Assessment, language education testing, washback effect

### 1. Introduction

Every kind of action in the nature is pregnant to a result, which is similar to reason and result circle. The same phenomenon keeps its existence in the field of education. Education is constructed as a process, and is need of assessment to observe and evaluate the process results. Assessment comes into the stage in many forms ranging from summative assessment to placement assessment. No matter what kinds of form the assessment is structured, the focal point is the ultimate aim and the result of it from the test-takers perspective. The effect of the examinations on the test takers decides the success level of the students since examination style is not the matter of the issue. The exact matter is the process of the test formulation and conduction. Hence, the test formulation, test conduction, and test evolution are the major characters together with the education process. It is an unneglectable fact that teaching and learning process are inseparable from the assessment whether it is process-oriented or product-oriented. By depending on the very-well interrelated fact of the education and assessment feature, the current review study was set to review the related literature critically from the perspective of the effects of the assessment of English language and its probable effects on the test-takers.

The rationale of the present study is the originality of reviewing the literature critically from the dimension of washback effect and English language assessment via central national exams employed in Türkiye within years in comparison to the pretty close similar assessment types in Japan and India. This review aims to present a critical review of literature regarding washback effect and its link with English Language and Testing. At the first step: history of washback effect is touched, and then the most common definitions of washback effect, positive washback – negative washback, washback effect and teacher relation, and widespread interiorized washback models are all critically reviewed with the purpose of comparing Türkiye's English Language testing with similar countries testing styles regarding the washback effect of testing on learning process. The following research questions has shed light on the research concerns.

1.What is the washback effect of the central language tests on test-takers?

2. What are the similarities and nuances between Türkiye’s language testing system and Japan’s and India’s testing system?
3. Are there any differences between the washback effects of the production-oriented exams and process-oriented exams?

## 2. Methodology

With the aim of reviewing and criticizing the related materials as well as presenting a fresh perspective based on earlier research, the current study begins with a narrative literature review (Snyder, 2019). Within this aim, primary topics (i.e.: Booth, 2012; Shih, 2007; Tsang & Isaacs, 2022; Zhan & Andrews, 2014) are analysed with regard to washback effect and language education and secondary headlines were also examined in order to have an intact understanding about the evolution of the washback effect in the history. This detailed analysis on the procedural link between washback effect and language education has enabled to bring a shared understanding on the particular washback effect of the central exams, in particular central foreign language exams, in Türkiye on test-takers. In the present review study all the ethical considerations are cared by strictly obeying the APA 7 citation rules both in text citations and references.

## 3. Findings

### 3.1. What is Washback Effect?

There are various terms used for expressing the effects of tests on learners, teachers, and stakeholders. Although the most common term is washback: ‘*backwash*’, ‘*test feedback*’, ‘*measurement-driven instruction*’, ‘*curriculum alignment*’, and *test impact*’ are the other terms addressing the same meaning as washback. Washback effect was analysed both in general multi-disciplinary education fields and language education field. Whilst definition of the term of ‘washback’ is context specific (Şentürk, 2013), it has been mostly uttered as the impact of testing on teaching and learning (Cheng, 2000; Katagiri, 2023, Pan, 2022; Marsden et al., 2023). Depending on the adopted context, ‘*test-impact*’ (Bachman & Palmer, 1996; Baker, 1991) and ‘*backwash*’ (Alderson & Wall, 1993; Buck, 1988) are among the common names used for washback effect. In the later years of the 1950s and in the beginning of the 1960s, the influence of exams on teachers, students, and stakeholders caught the attention of researchers from the core studies (Hungerland, 2004). And external exams which are also called as high-stake or central exams are the protagonists of these test-driven classes as their effect is assumed to be higher than the local classroom-sized exams (Hungerland, 2004; Pan, 2022). Washback effect is categorized as macro-level and micro-level: central exams and high-stake exams are counted as macro-level exams, and local (school-based) exams are counted as micro-level exams. So it is easy to grasp that exams’ categories are named according to their impact on the target audience.

### 3.2. Is Washback Effect positive or negative?

Depending on the effect of the test on the target audience, washback effect is classified as positive washback or negative washback. Brown (1995) simplifies the situation as following; if a test meets the requirements of the syllabus, it has positive washback; if a test fails to meet the requirement of the course content and cause a deviation, it has negative washback. So it is possible to conclude that washback can be either harm or benefit in the field of language education. Brown (2002) tries to sum up both harmful and beneficial washback in a diagram:

<i>Negative washback affect:</i>	<i>can</i>	<i>Positive washback fostered:</i>	<i>can</i>	<i>be</i>
<ul style="list-style-type: none"> <li>Teaching</li> <li>Course content</li> </ul>		<ul style="list-style-type: none"> <li>Alter test design factors</li> <li>Change test content factors</li> </ul>		

- 
- |                          |                                      |
|--------------------------|--------------------------------------|
| ● Course Characteristics | ● Adjust test logistics factors      |
| ● Class Time             | ● Modify test interpretation factors |
- 

*Figure 1. Negative and positive washback by Brown (2002).*

### 3.3. Positive Washback

Pearson (1988) asserts that well-designed tests are made use of and shared as the beneficial course material since they ensure the positive washback effect. So, what is positive washback? If the assessment criteria meet the requirement of the content, it is possible to mention a positive washback effect there (Brown and Hudson, 1998). With basic terms, if the effect of the teaching and testing process all together leaves a positive effect on the target audience, it is called as positive washback. For instance, if course objectives in a syllabus is set in order to assess students' communicative abilities and observe and test students' employing communicative tasks (e.g. presentation, storytelling), an effective positive washback effect can be traced: since the objectives and assessment criteria are parallel.

---

<b>Positive Washback</b>	
<hr/>	
<b>Classroom settings</b>	1. Tests induce teachers to cover their subjects more thoroughly, making them complete their syllabus within the prescribed time limits
	2. Tests motivate students to work harder to have a sense of accomplishment and thus enhance learning.
	3. Good tests can be utilized and designed as beneficial teaching-learning activities to encourage positive teaching-learning processes.
<b>Educational/ societal system</b>	1. Decision-makers use the authority power of high-stakes testing to achieve the goals of teaching and learning, such as the introduction of new textbooks and new curricula

---

**Figure 2.** Summary of the Positive Washback (Pan, 2009, p.261).

In the context of Canada, Turner (2001) states that if the learners (test-takers) are included into test development process, positive washback is promised. Depending on this, Saif (2006) made a research and in this research tried to explore whether Turner's claim work or not, he involved the students to test preparation step and the result was positive in a Canadian university. Muñoz and Álvarez (2010) also reported the positive result in terms of positive washback and learner involvement in relation to the test development process in the following years. Whilst it is stated that including and reshaping teaching methods, course design, and assessment criteria is effective in positive washback (Alderson & Hamp-Lyons, 1996; Alderson & Wall, 1993a, 1993b; Gates, 1995; Katagiri, 2023; Marsden, 2023; Messick, 1996; Pan, 2009; Pan, 2023; Shohamy, 1993; Şentürk, 2013), it was proven

that involving learners to the test development step guarantees the optimal positive washback. (Muñoz & Álvarez, 2010; Saif, 2006; Turner, 2001).

In addition to the aforementioned characteristics to ensure the positive washback, Brown (2000a) has drafted beneficial strategies under four categories: test design strategies, test content strategies, logistical strategies, and interpretational strategies.

#### ***A. Test design strategies***

1. Reach a high number and fluctuating sample. Employ direct testing methods.
7. Ensure autonomous learning activities.

#### ***B. Test content strategies***

1. Focus on the desired abilities.
2. Focus on productive items more than multiple-choice tests.
3. Include all areas of the curriculum.
4. In order to ensure cognitive skills are taught, concentrate on them.
5. Test all four skills.
6. Do not restrict the test items to academic skills, include also basic skills.
7. Adopt real-life tasks.

#### ***C. Logistical strategies***

1. Ensure that all target audiences are aware of the aims of the test.
2. Make all the goals clear and focussed on the main areas.
3. Provide teachers with assistance about the tests where necessary.
4. Present immediate feedback to teachers and ensure the exchange of the information.
5. Provide prompt and due feedback to teachers and schools about the difficulty level of the test and student's competence level in national examinations.
6. Include all the administrators and teachers to the application process of the tests.
7. Inform teachers and schools back with precise score reports.

#### ***D. Interpretation strategies***

1. The results of the exams should be trustable, credible, and fair for test takers and score user institutions.
2. Keep in mind the elements other than the teaching process in the evaluation steps.
3. Employ foretelling validity studies of national examinations.
4. Basically, for the test design step, develop well-prepared competence of testing authorities.
5. Beware that every research board has its own working capacity.
6. The Curriculum board and the testing board should be in close relation.
7. Build well-designed networks to foster exchange of information in the same region.

The blanket style (covering all of the crucial factors) of this strategy list makes it valuable and useful. Every step of test preparation is mentioned in details in the lists. Moreover, as the order of the strategies

in chronological order, it can be grasped as pre-testing, in-testing, and post-testing steps (Dağıtan, 2015).

### **3.4. Negative washback effect**

It is commonly accepted that negative washback is the result of recognition centred tests which are generally multiple-choice tests. However; some scholars agree that it is the natural consequence of misusing tests, not the style or type of the tests (Alderson & Wall, 1993; Heyneman, 1987; Katagiri, 2023; Marsden et al., 2023; Messick, 1996; Pan, 2022; Pearson, 1988). The negative washback effect is more common than the positive washback effect in the field of education (Dağıtan, 2015). The most accepted expression of negative washback effect is that if the course syllabus does not match the content of the tests or meet the requirements of the concerned courses, this will effect learners' studying style natively, which is also called as Negative washback. The most widespread case is that learners are motivated to get high marks from exams, and as a consequence, they have a tendency to study only exam-based content and ignore the rest of the syllabus adapted in the classrooms. In turn of this situation, learners, parents, administrators, and stakeholders concern only the exam-oriented educated style and consider life skills, and higher-order skills are perceived as unnecessary skills to include into the curriculum.

### **3.5. How to overcome negative washback effect?**

As it is not possible to assess somebody without any evaluation criteria, and getting rid of any type of exams, it is logical to accept the situation and try to search for ways to decrease its negative washback effect as much as possible. High-stake exams are shown as the main reason for the negative washback effect, and it is claimed that conducting exams to a higher number of persons is only possible with high-stake exams (Alderson & Wall 1993; Katagiri, 2023; Marsden et al., 2023; Pan, 2022; Shohamy 1993, 2001; Smith, 1991; Wall 2000). Since the main consideration here is the economy of time, effort, reliability, and validity for the accountability of the exams. Kiany, Shayestefar, Samar, and Akbari (2006) state that high-stake exams are chosen as a remedy for promoting the activity of learning and materials and making them valid; but, this target has been ended with failure in many contexts as it does not ensure the long term goals. Policy-makers try to get precautious against negative washback effect through ensuring the perfect match between course content and exams. This is only one of the precautions in general term and it does not guarantee to omit the negative washback effect totally, but minimizing it as much as possible. Precaution types are shaped from country to country and from lesson to lesson as it is domain-specific. And Kiany et al. (2006) state that precautions are generally taken after the implementation and observation of the process.

Policy-makers stress that high-stake exams have a gatekeeping mission in many institutions, so it is natural to have some sever consequential effects along with them (Alderson & Wall, 1993; Shohamy, 1993, 2001; Smith, 1991; Wall 2000). Regarding this situation, operative reforms which are rote-learning, lower-level skills, and touching to authentic, and real-life skills to overcome the negative washback effects are among the suggestions to lower the negative washback effect. Nevertheless, Kiany, et al., stress that empirical studies have shown that high-stake exams have fallen in short of reaching the target of positive washback effect.

### **3.6. Teachers' Role in Washback Effect**

If education is thought as a circle of a chain, then teachers are the first circle of this chain since they are the jumping place of any movement (Orak, 2019). Also, as aforementioned in under the title of CLT, teachers' role overweighs the other variables such as learners, materials, and curriculum (Brown, 2000; Katagiri, 2023). Since classrooms are the stages of the teachers, teachers may decrease the washback effect minimum if they collaborate with learners and take responsibility to promote CLT related tasks even if they are not stated openly in the curriculum (Bachman & Palmer, 1996). CLT values interaction more than linguistic rules, so teachers may organize their course content and exam



contents accordingly to ensure the rapport between them. As a result learners will get the message that interaction brings language acquisition rather than pure grammar and translation filled with courses (Booth, 2012; Tsang & Isaacs, 2022; Zhan & Andrews, 2014).

While teachers may change many things in the classroom (Alderson & Hamp Lyons, 1998; Cheng, 1997, 2000; Katagiri, 2023; Pan, 2022; Shohamy, 1994, 2004; Turner, 2001, 2006, 2009; Wang, 2010), they are tied hand and foot. Owing to social, cultural, political, and administrative barriers, teachers are not free to set their own language education courses and develop their exam questions. Even though school based-exams are prepared by teachers, learners will sooner or later sit the national high-stake exams. Since learners aware of the situation, they are motivated to learn only high-stake exam-related subjects or course content. This is the same case in Japan, teachers ignore the CLT related tasks in language education as these tasks do not prepare the learners to the nationwide exams (Gorsuch, 2000, 2001; Nishino & Watanabe, 2008; Taguchi, 2005).

Furthermore, Akiyama's (2004) and Taguchi (2005) underline in their researches that test design is also effected by washback phenomenon, speaking skills are not tested. The same situation goes for Türkiye in Nationwide language exams (YDS and YÖKDİL). As speaking skill is not tested in the language exam, language courses do not focus on the courses including speaking tasks. Another fundamental thing is teachers' cultural and educational backgrounds, working stories, their characteristics, and professional sophistication. Those background elements should be examined closely apart from other variables.

### **3.7. Washback Effect and Exam Types**

There is a close relationship between the exam types and eventual washback effect of them on test-drivers and test-takers in addition to policy-makers and stakeholders. The washback effect of the exam types on these parts diverges noticeably when the teaching activities, learning activities, time, effort, and economical concerns are taken into account. Recognition-oriented and production-oriented exam types shape the discussion on the washback effect and exam types; plus and minus shades of the exam types are uttered together with a stick with two ends. There is the concern of teaching and learning at the one end, and the concern of effort, economical issues, and time at the other end. These concerns are going to be reviewed in detail under the following subtitle.

### **3.8. Production-oriented exams versus proses-oriented exams**

Recognition-oriented exam types are named as production-oriented exam types as well, they are referred reciprocally. When the recognition-oriented exam type is the matter of issue, standardized tests are the first ones come into mind: since, they are employed to a large extend globally. Standardized tests are mainly preferred at high-stake exams in the form of multiple-choice. Multiple-choice exams do not require any forms of production, it rings the bell for the recognition by presenting choices to the test-takers. When recognition-oriented multiple-choice tests are evaluated from the perspective of test-takers, it has both pros and cons. From calling back the pieces from the whole in the immediate exam environment: so, multiple choice exams are sympathized highly by test-takers though it is not a favour for them. In addition to helping to remember through choices, it also brings the chance factor to the stage. Even if the test-takers do not remember or know anything about the questions, first they may eliminate some irrelevant choices, and decrease the number of choices through recalling their similar knowledge. And after decreasing the number of choices to two, they may try their chance by selecting either of them. Apart from test-takers, test-makers and test-drivers also benefit from the pros of the recognition-oriented exams in terms of preparing step, administering step, and evaluation step. Recognition oriented exams run well will high number of populations: since it expedites all of the exam stages ranging from preparing to evaluation. When preparing the recognition-oriented exams, test-designer do not think about the variety for each test-taker as all of the test-takers are all going to respond to the same questions, may be just in different order. Furthermore,

it is convenient to secure validity and reliability in recognition-oriented exams, especially for the multiple-choice ones. Secondly, administration of the recognition-oriented exams is powerful in terms of economy of time, staff, and effort at the administration stage. It makes administering the exam possible in a short time, with little effort, and by employing a few numbers of staff. Finally, it is also attractive in terms of evaluation, mostly for the subjectivity concern. Recognition-oriented exams do not require human interpretation, it is mainly mechanic, which brings the convenience of objectivity of the results. Moreover: since evaluation is mechanic, results are announced in a short time after the exam. As a natural result of the aforementioned pros, recognition-oriented exams are highly supported by the states to be employed by the Ministry of the Higher Education and Council of Higher Education.

While on the one hand recognition-oriented exams are highly favoured in terms of administration and evaluation, on the other hand, it is criticized widely cause of the washback effect of them on teachers and learners. With more specific examples, in Japan, Taiwan, and Korea learners are only trying to study the exam-related subjects, not the subjects related to basic life skills. In America, high school students are focusing on only the SAT exam (Marsden et al., 2023; Pan, 2022; Brown, 2011), and in Türkiye, main aim of the high school seniors is to pass the university entrance exams (LGS, YGS). When we evaluate the situation from the point of learning and teaching a language, it is going to be clearer to grasp the severity of the situation. Recognition-oriented exams direct both teachers and learners to teaching and learning grammar more than four basic skills (reading, writing, speaking, and listening). While learning a language requires basic life-skills, washback effect of the recognition-oriented exams shapes the priorities of the teaching and learning process.

### **Primary recognition-oriented exams in Türkiye**

When the recognition oriented exams are the matter of discussion, generally multiple-choice forms are the first ones come into mind: since they are expansively adopted in high stake exams. And administration of high stake exams is very common in Türkiye like all highly populated countries around the world. While objectivity and practicality are claimed to be basic reasons of the widespread adaptation of high-stake exams in the form of multiple-choice tests, the root reason stands as the economy of preparation and administration of the exams at the blind part of the iceberg.

In Türkiye, high-stake exams start at the Secondary School for the entrance to High Schools, and it is called as High School Entrance Exam (LGS-Turkish abbreviation). After 4 years of high school education, students sit another high-stake exam for university entrance exam called as Entrance Exam to Universities, held in two days with three different sessions. There are also high-stake exams in order to transfer from Associate Degree programmes to Bachelor Degree. After university education, students need to sit Public Personnel Selection Examination (KPSS- Turkish abbreviation) if they want to be employed by public institutions. In terms of testing English proficiency, there are high-stake exams for testing English proficiency called as Foreign Language Exam (YDS) and Language Exam of Higher Education Council (YÖKDİL). English proficiency high-stake exams get most of the criticism upon their nature of purportedly spiritless, automated, and recitative form, which is submerged with grammar and vocabulary, both parties- the ones who can pass or not-. Another remarkable criticism is that English proficiency high-stake exams are far from the requirements of linguistic standards for academicians, since academicians need the language for their profession. They need to use it in their daily life to produce science, not only for proving their English proficiency on paper. By relying on these natural and unsurprising criticisms, English proficiency high stake exams fall behind meeting the needs of the test takers not only in Türkiye but also all over the world. It should not be forgotten that if the assessment is for teaching, than the motto should be 'Language is for life'. This motto even demonstrates the exact vision that should be adopted in language education, and it helps to decrease the washback effect of high-stake exams to a certain degree.

Apart from those, there is one another high stake exam in the name of Academic Personnel and Postgraduate Education Entrance Exam (ALES) employed to academic candidates who are in seek of

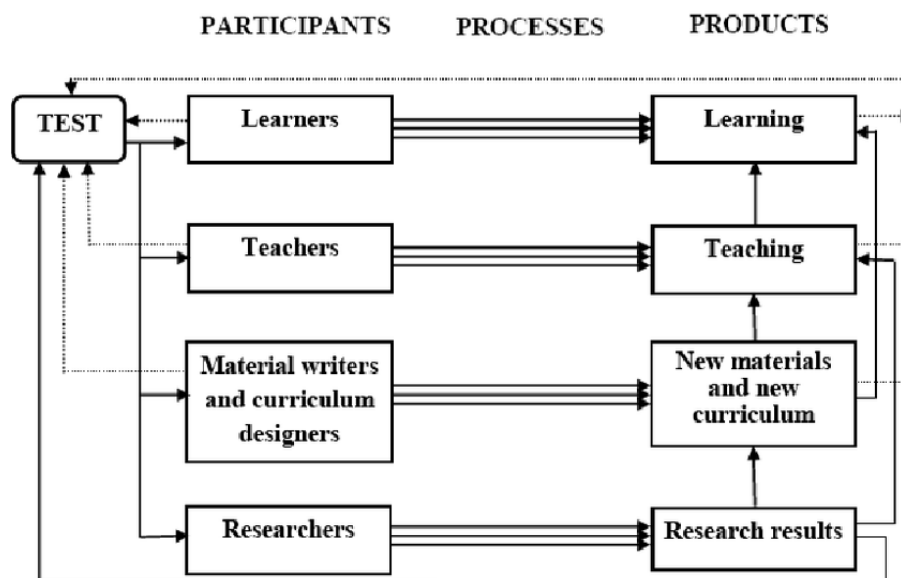
promotion, aiming to be employed as lecturer by universities (private and public), or aiming to start master degree or doctorate degree in Türkiye. It is conducted in Turkish and comprises only Maths and Turkish Language tests within a really limited time with long questions.

### 3.9. Washback Models around the World

While washback effect and its positive and negative consequences are under discussion, scholars tried to summarize all of the characteristic elements consisting of washback effect in a structured frame basing on their approach to washback effect theory. When it is accepted as a framework, washback effect includes interrelated elements constituting the final picture (Dağıtan, 2015).

#### 3.9.1. Bailey washback effect model

Bailey posed a model in 1999 to illustrate the multi-dimensional way among the elements and when these elements are evaluated holistically, washback effect is seen at the top of this frame. In the model of Bailey, there is an interwoven relationship between the final product and process which reciprocally issues from and revisit the eventual destination which is ‘TEST’ (see Figure 3).

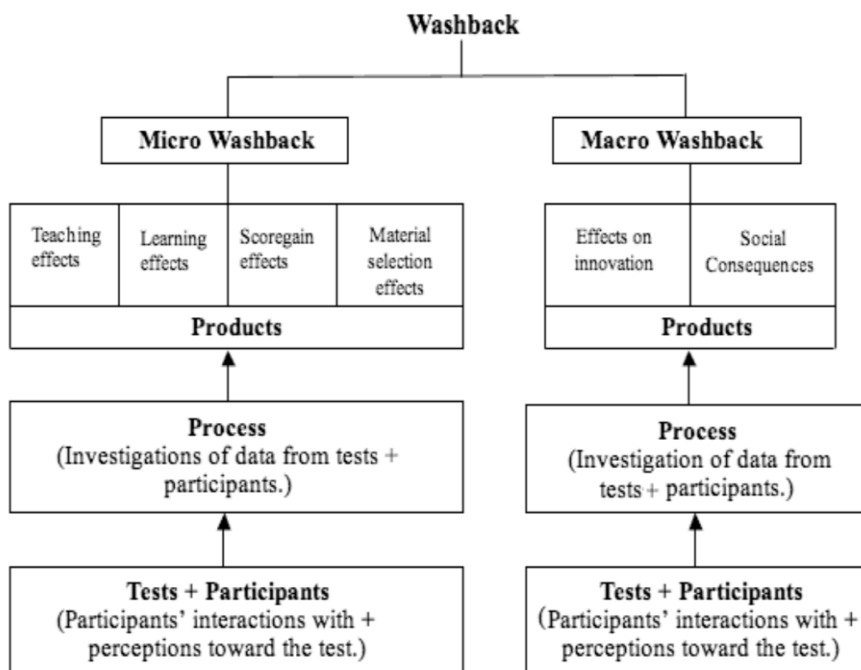


*Figure 3.* Washback Model by Bailey (1999, p. 11)

#### 3.9.2. Pan washback effect model

Impressed by the holistic ideas of Bachman and Palmer (1996), and Hughes (1993) on washback effect, a washback model based on two-axis frame was illustrated by Pan in 2008. In this model, Pan depicts fixed borders between micro and macro levels, and through this borders Pan clarified public and individual effect size recognizably. Pan stresses that in order to analyse the effect size, triangulation should be ensured by observations, interviews, questionnaire, pre-tests, and post-tests. In his model, Pan draws attention to not only test takers but also stakeholders, administrators, policy makers and family members (see Figure 4).

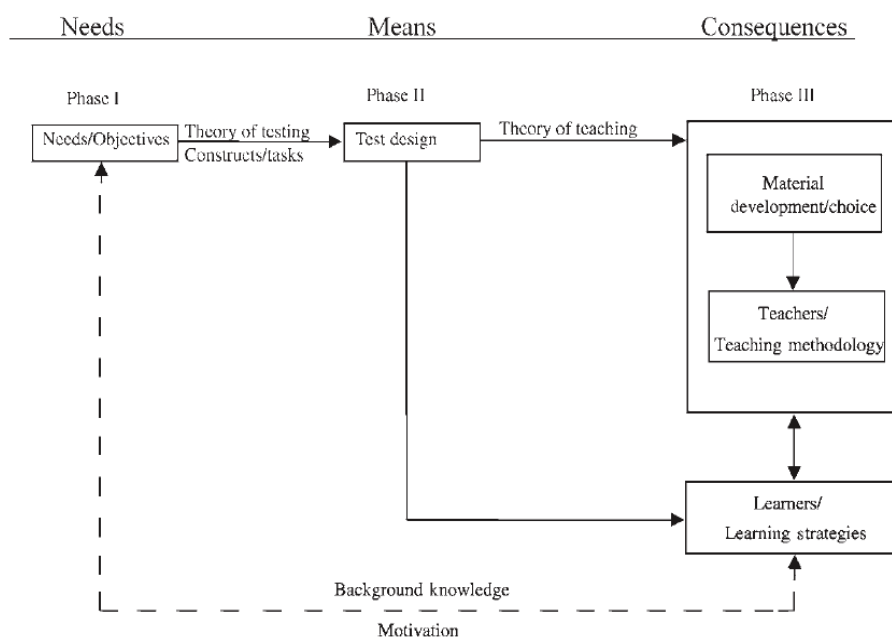




**Figure 4.** Holistic Washback Effect Model by Pan (2008, p.12)

### 3.9.3. Saif washback effect model

In 2006, another washback model was portrayed by Saif, it is similar yet very basic when compared to the aforementioned models. In this model, background knowledge and motivation are depicted as the vital elements. In contrast to the models depicted by Bailey (1999) and Pan (2008), Saif (2006) illustrated the whole process in three-steps (respectively); *needs*, *means*, and *consequences* (see Figure 5).



**Figure 5.** A Conceptual Washback Effect Model by Saif (2006, p.5)

## 4. Discussion & Conclusion

Discussion of the findings is given in a thematic presentation.

#### **4.1. Testing of English in Türkiye and Its Washback Effect**

It is generally agreed today that learning a foreign language and being able to use it effectively has become a must as international business affairs have gained significant importance because of the globalization and going abroad for the education has become popular. In addition to these factors, speaking a foreign language has been a prestige among the members of the society. While the learning process takes place, the learners have to sit for various tests to show how much they have learnt or to see how their progress is. That is why tests or examinations have gained important roles in our educational system.

In Türkiye, students at private schools start learning English at a very early age like 5 or 6 and the learning process takes place till they finish high-school at the age of 18. But in the state schools, this process starts at the age of 10 when they are 4th graders and finishes by graduating from the high school at the age of 18. There is also a big difference between the numbers of English lessons studied at the state schools and private schools. Since teaching a foreign language is one of the major aims at the private schools, the numbers of English lessons show intensity and the number of the lessons may change between 8 lessons and 12 lessons in a week. When we look at the state schools, the students have 4 hours of English lessons in week and this number is really inadequate for the students to learn a foreign language. The students at the private schools have more chances to sit for an international test while the state school students do not unless they prefer studying foreign languages at a university.

#### **4.2. Main Criticism to High-Stake Exams in Türkiye**

High-Stake exams which can be counted as university entrance exam, High School Entrance Exam, National Language Exams, are mostly criticized from the perspective of their being recognition-based tests and focusing on highly mechanical skills. It is stated that lesson contents are not at the centre of the learning process as a lifelong skill for long term goals but as a mechanical skill for short term goals. As a natural outcome of this perspective, learning does not take part in this educational activity, the aim is only to pass the exams (Dağitan, 2015; Suen & Yu, 2006). Among the high-stake exams, national language exams (YDS- Foreign Language Exam, YÖKDİL- Council of Higher Education Foreign Language Exam) are at the centre of criticisms, since while language exams should be interactive and production-centred, but they are conducted in recognition-based style for example multiple-choice exams. It is underlined that mechanic exams only focuses on linguistic skills not real life language. Test-takers complain that they only learn about the grammar rules and learn by heard the vocabulary. However, after getting the sufficient score they just forget most of the rules and they cannot apply the rules to their daily life needs or academic life needs. So, by depending on the needs, it is possible to conclude that high-stake exams do not meet the addressed needs of the test-takers which can be also considered as a negative washback effect of the high-stake exams, and it is said that that kind of high-stake exams are considered to be as ‘underrepresent’ (Katagiri, 2023; Marsden et al., 2023; Suen & Yu, 2006).

#### **4.3. Comparison of Türkiye, India, and Japan Cases in terms of Testing and Washback Effect**

When the education systems are compared, it is seen that Japan’s and India’s education system are very similar to Turkish education system in terms of testing and schooling procedures. Therefore, they were selected as the pivotal examples to discuss and criticize the washback effect of the testing regarding the case of Türkiye. In all these three countries, obligatory education starts from kindergarten at the age of 5 and continues with primary school, and ends with high school (vocational and technical high school apart from Anatolian high schools and super high schools in Turkish context). University education (B.A.), master degree (M.A.) education and doctoral degree education (Ph.D.) depends on the individuals’ choices. High number of the population in these countries is shown as the reason of employing high-stake exams for high school education and university degree. While the primary school and high school education are obligatory, university education depends on the

individuals' preference, but if individuals choose to go on their education at university degree, governments support them with scholarships and dormitory facilities.

Another common point in these three countries is private courses and weekend courses. Private courses and weekend courses are adopted by families as a tool to provide students with additional materials and lessons by teachers on their school week subjects. Families use every means of supportive sources for their children as competition mood is really active in these countries: since population is high and job opportunities are scarce. That is why job opportunities are considered as directly related to high-stake exams and its active washback effect on learners. With another term, washback effect is the reflection of the states' life conditions on their education policies.

Although these three countries seem quite similar in terms of examination system as the natural outcome of the growing population, Türkiye is more different from them. While Japan and India have promising outcome from the higher education institutions regardless of the population, Türkiye falls behind them. The most probable reason seems that the colonial history behind the roots. Colonial history of India explains that though India has high-stake exams, citizens are pretty good at speaking English since they have gained the precious conscious of learning another language which is English in their case. Japan's citizens are good at English, as well and it is explained with the high rate of trade in their land ranging from technology to industries. In the case Türkiye, apart from the negative washback effect of the high-stake exams, national language education policy has not been improved in parallel to technology and 21<sup>st</sup> century need. It still tackles with the high number population and high-stake exams with the low-level of foreign language conscious.

History of washback effect of exams on test-takers and test-makers dates back to old times, but it started gaining popularity in 1993 when Alderson and Wall (1993) made seminal research on it. From that time on there have been plethora of studies on the relation between washback effect and language education. Hungerland (2004) cites that there are many hypotheses based on the hypothesis explained by Alderson and Wall in 1993, and those hypotheses have all been supported by various researchers administered to various participants. Shohamy searched washback effect in Israel context from ELT and testing perspective in 1996, and reported that tests are effective tools for directing the learners' behaviours and even attitudes towards ELT. Again, in the same year Alderson and Hamp-Lyons searched the washback effect of TOEFL exam, and Watanabe did similar researches in 1996 and 2000 in Japan contexts. All of them reported a significant washback effect of exams on both learning and teaching activities. In line with the aforementioned statements, it is possible to deduce that there is a direct proportion between washback effect and high-stake exams in the negative direction. High-stake exams accelerate the washback effect in the direction of mechanical learning and teaching activities. This brings the employment of the repetitive-cyclical teaching methods: both test-takers and test-makers follow the same circle as the eventual aim is to pass the exam, not learn the language. Which means that all learning and teaching activities are under the tomb of exams. Educational programs and in line with them: course content is planned in accordance with the content of the exams. That is why it is not logical to expect real-life skills from the victims of this systems.

On the other hand, it is an unfair and pessimist argument to claim that washback effect happens only in the negative direction: since washback effect of exams in the field of ELT is also encountered in the positive direction. Thanks to adoption of alternative assessment tools in language education, the perceived negative washback effect of exams on test-takers and test-makers is directed into positive and more promising language activity. It is highly recommended that testing should be for learning, and ultimate purpose must be assessing the whole process, not focusing on the exam results (Brown, 2000; Katagiri, 2023, Marsden et al., 2023; Pan, 2022). As the role of the teachers in the language education is like being the protagonist in a play, teachers (test-makers) have the power to shape the washback effect on learners (test-takers).

One another the substantial fact is that 21<sup>st</sup> century global world is expecting graduates armed with real-life skills more than graduates armed with mechanic skills or graduates fulfilled with high scores. It means that theoretical knowledge should be integrated with 21<sup>st</sup> century learning and innovation skills - 4Cs (communication, critical thinking, creativity, and collaboration). If language education is left to the arms of high-stake exams and in turn washback effect in ELT activities, ELT will sooner or later fall behind the requirements of the century. So, washback effect should be benefitted by converting it into positive way in order to meet the addressed needs of the language education emerged from the immediate century.

### Conflict of interest

The author hereby declares that there is no conflict of interest in the present study. There is a single author for the present review study. The author confirms that there is no supporting institution to complete the study. All the ethical considerations are cared by obeying the APA 6<sup>th</sup> citation rules in a review study.

### References

- Akiyama, T. (2004). *Introducing EFL speaking tests into Japanese senior high school entrance examination*. [Unpublished PhD thesis, University of Melbourne]. ProQuest
- Alderson, J. C., & Hamp-Lyons, L. (1996). TOEFL preparation courses: Washback. *Language Testing*, 13, 280-297.
- Alderson, J. C., & Wall, D. (1993a). Does washback exist? *Applied Linguistics*, 14, 115-129.
- Alderson, J. C., & Wall, D. (1993b). Examining washback: The Sri Lankan study. *Language Testing*, 10, 41-69.
- Bachman, F. & Palmer, S. (1996). *Language testing in Practice*, Oxford University Press.
- Bailey, K. M. (1999). Washback in language testing [TOEFL Monograph Series]. Princeton: Educational Testing Service.
- Baker, E. (1991, May 5-7). Alternative assessment and national policy. [Conference session]. *1<sup>st</sup> National Research Symposium on Limited English Proficient Students' Issues: Focus on Evaluation and Measurement*, Washington, DC, USA.
- Booth, D. K. (2012). Exploring the washback of the TOEIC in South Korea: A sociocultural perspective on student test activity. Ph.D. dissertation, The University of Auckland.
- Brown, J. D. (2000). *What is construct validity?* Longman.
- Brown, J. D. (2000a). University entrance examinations: strategies for creating positive washback on English language teaching in Japan. *Shiken: JALT Testing & Evaluation SIG Newsletter*, 3(2), 2-7.
- Brown, J. D. (2002). Extraneous variables and the washback effect. *Shiken: JALT Testing & Evaluation SIG Newsletter*, 6(2), 10-13.
- Brown, J. D., & Hudson, T. (1998). The alternatives in language assessment. *TESOL quarterly*, 32(4), 653-675.
- Buck, G. (1988). Testing listening comprehension in Japanese university entrance examinations. *JALT Journal*, 10(1), 15-42.
- Cheng, L. (2000). *Washback or backwash* (Report No. FL 026 284). Kingston: Queen's University.

- Cheng, L. (1997). How does washback influence teaching? Implications for Hong Kong. *Language and Education*, 11, 38–54.
- Dağtan, E. (2015). *Üniversitelerarası Kurul Yabancı Dil Sınavı'nın geri etki (Washback) açısından incelenmesi: Dicle Üniversitesi örneği* [An examination on the Washback effect of the Interuniversity Board Foreign Language Test: Dicle University Case, Master thesis, Dicle University-Diyarbakır] YÖKTEZ.
- Gates, S. (1995). Exploiting washback from standardized tests. In J. D. Brown S. O. Yamashita (Eds.), *Language testing in Japan* (pp. 101-106). Japan Association for Language Teaching.
- Gorsuch, G. (2000). EFL educational policies and educational cultures: Influences on teachers' approval of communicative activities. *TESOL Quarterly*, 34(4), 675-710.
- Gorsuch, G. (2001). Japanese EFL teachers' perceptions of communicative, audiolingual and *yakudoku* activities: The plan versus the reality. *Educational Policy Analysis Archives*, 9(10).
- Heyneman, S. P. (1987). Use of examinations in developing countries: Selection, research, and education sector management. *International Journal of Education Development*, 7, 251–263.
- Hughes, A. (1993). *Backwash and TOEFL 2000*. University of Reading.
- Hughes, A. (1989) *Testing for Language Teachers*. Cambridge University Press.
- Hungerland, R. (2004, January 27-29). Language proficiency testing: A critical survey. [Conference Session]. *TESOL Canada National Conference, Canada, North America*.
- Katagiri, W.N. (2023). How sociocultural factors mediate washback effect: a study of the Hong Kong Diploma of Secondary Education Examination English Language Paper 4. *Lang Test Asia*, 13(1), 1-18.
- Marsden, E., Dudley, A., & Hawkes, R. (2023). Use of word lists in a high-stakes, low-exposure context: Topic-driven or frequency-informed. *The Modern Language Journal*, 107, 669–692.
- Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13, 241-256.
- Muñoz, A. & Álvarez, M. (2010). Washback of an oral assessment system in the EFL classroom. *Language Testing*, 27(1), 33-49.
- Nishino, T., & Watanabe, M. (2008). Communication-oriented policies versus classroom realities in Japan. *TESOL Quarterly*, 42(1), 133-138.
- Noble, A. J., & Smith, M. L. (1994a). *Measurement-driven reform: Research on policy, practice, repercussion* (Tech. Rep. 381). Arizona State University, Center for the Study of Evaluation.
- Pan, YC. (2022). Teacher and Learner Washback of a High-stakes Lexical Grammar Writing Test. *The Journal of Asia TEFL*, 19(3), 977-996.
- Pan, Y. (2008). A critical review of five language washback studies from 1995-2007: Methodological considerations. *Shiken: JALT Testing & Evaluation SIG Newsletter*, 12(2), 2-16.
- Pearson, I. (1988). Tests as levers for change, in Chamberlain, D. and Baumgardner, R. J. (Eds) *ESP in the classroom: Practice and evaluation* (pp.98-107). Modern English Publications.
- Saif, S. (2006). Aiming for positive washback: a case study of international teaching assistants. *Language testing*, 23(1), 1-34. doi: 10.1191/0265532206lt322oa
- Shih, C. M. (2007). A new washback model of students' learning. *The Canadian Modern Language Review*, 64(1), 135–162.



- Shohamy, E. (1993). The power of tests: The impact of language tests on teaching and learning. *NFLC Occasional Paper*. Washington, DC: National Foreign Language Center.
- Shohamy, E. (2004). The power of tests over teachers: The power of teachers over tests. In D. J. Tedick (Ed.), *Second language teacher education* (pp. 101-111). Lawrence Erlbaum.
- Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, 104, 333-339. <https://doi.org/10.1016/j.jbusres.2019.07.039>
- Suen, L. K., & Yu L. (2006). Chronic consequences of high-stakes testing? Lessons from the Chinese civil service exam. *Comparative Education Review*, 50(1), 46-65.
- Taguchi, N. (2005). The communicative approach in Japanese secondary schools: Teachers' perceptions and practice. *The Language Teacher*, 29(3), 3-9.
- Tsang, C. L., & Isaacs, T. (2022). Hong Kong secondary students' perspectives on selecting test difficulty level and learner washback: Effects of a graded approach to assessment. *Language Testing*, 39(2), 212-238.
- Turner, C. (2001). The need for impact studies of L2 performance testing and rating: identifying areas of potential consequences at all levels of the testing cycle. In A. Brown et al. (Eds.), *Experimenting with uncertainty: Essays in honour of Alan Davies, studies in language testing 11* (pp. 138-149). Cambridge University Press.
- Turner, C. E. (2006). Professionalism and high-stakes tests: Teachers' perspectives when dealing with educational change introduced through provincial exams. *TESL Canada Journal*, 54-76.
- Turner, C. E. (2009). Examining washback in second language education contexts: a high stakes provincial exam and the teacher factor in classroom practice in Quebec secondary schools. *International Journal of Pedagogies and Learning*, 5, 103-123.
- Wang, J. (2010). *A study of the role of the 'teacher factor' in washback*. [Unpublished doctoral dissertation, McGill University]. Montreal.
- Wiseman, S. (1961). The efficiency of examinations. In S. Wiseman (Ed.), *Examinations and English education*. Manchester University Press.
- Zhan, Y., & Andrews, S. (2014). Washback effects from a high-stakes examination on out-of-class English learning: Insights from possible self -theories. *Assessment in Education: Principles, Policy & Practice*, 21(1), 71-89.