

Washback Effect of TOEIC Listening And Reading as a College Exit Test in Riau, Indonesia

Masrul¹⁾, Ummi Rasyidah²⁾

¹⁾ Universitas Pahlawan Tuanku Tambusai, Indonesia

²⁾ Universitas Pasir Pengaraian, Indonesia

SUBMISSION TRACK

Submitted : 28 October 2023
Accepted : 20 November 2023
Published : 25 November 2023

KEYWORDS

TOEIC, listening, reading, test impact, standardized test.

CORRESPONDENCE

E-mail: masrulm25@gmail.com

A B S T R A C T

The utilization of the TOEIC test in some universities in Indonesia as a language exit exam highlights the phenomenon of the washback effect, revealing that this exit requirement influences the narrowing of teaching content and a less communicative teaching process. Some universities in Indonesia have employed the TOEIC (Test of English for International Communication) listening and reading test as a language exit exam. This study investigated teachers' perceptions concerning the TOEIC listening and reading test. The data were gathered by administering a questionnaire at three major Indonesian universities. The finding revealed that using TOEIC listening and reading as an exit requirement impacted the narrowing of teaching content and led to a non-communicative teaching process, yet the influence was not significant. This study found that the perceived test impact was influenced by factors differing from another research context.

Copyright © 2023 All right reserved

This is an open-access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license



Introduction

Many studies focus on the impact of English tests, including TOEIC. Some universities require students to take the TOEIC test as an exit test. The TOEIC test determines daily English abilities in a global work environment. Over the years, the demand for such tests has continued to increase English proficiency. The ability to speak English is often viewed as vital. It has been established that the TOEIC test results are directly proportional to the test taker's ability to perform English in the context of work and daily activities; the TOEIC test results have also been mapped into a broadly agreed system for explaining language proficiency (Powers et al., 2008; Booth, 2018; Park et al., 2020; Powers, 2018; Wilson, 2000). TOEIC test is divided into listening and speaking skills. Wilson (2000) measured the comprehension of English listening and reading comprehension skills of native Japanese and Korean speakers. To get a higher score, there is the TOEIC test, Powers et al. (2008) Facilitate the interpretation of test scores from the TOEIC test (listening and reading), which is redesigned as a measure of English proficiency, provides self-assessment inventories to TOEIC test takers that gather perceptions about their ability to perform various English tasks. The research found that participants who consistently performed daily English tasks scored higher (Powers et al., 2008).

Several studies have discovered a strong correlation between listening, reading, and overall English proficiency among students learning English as a second language (Jung, 2010). It was also stressed that listening comprehension is linked to reading comprehension and that a person's grammar and vocabulary levels are also linked to reading comprehension. As the TOEIC exam has increased in popularity, many universities have integrated TOEIC-based course instruction into their core curriculum. Park et al. (2020) focused on teaching Reading Comprehension (RC) and Listening

Comprehension (LC), assuming that reading and listening skills are interrelated, to improve students' TOEIC test scores. Many academicians agree that teaching TOEIC as part of the university curriculum is important. KI Shin & Oh (2012) examined the findings of the TOEIC test scores to emphasize the importance of TOEIC instruction. They also stressed that listening comprehension skills would help students obtain better TOEIC test scores in a shorter period. They believed TOEIC teaching successfully improved students' skills and produced almost immediate outcomes.

On the other hand, several studies investigated the challenges and solutions for the TOEIC test of students in junior high school. Zahrani et al. (2020) revealed that students encountered many problems, including their perceptions of vocabulary difficulties, grammatical difficulties, time constraints, and learning by grouping and searching the Internet. The TOEIC learning process is tailored to students' abilities. The TOEIC learning process continues by defining words, comprehending sentence structure, reading comprehension, and seeking to practice questions within a specified period. All of this is during the TOEIC exam, which will have a fixed time limit. As a result, this procedure must be followed for the forthcoming TOEIC exam to cover grades and how students can learn effectively and safely. Therefore, students are more motivated to learn and fully prepare for the TOEIC test. Furthermore, the research conducted by Mahaputri (2013) focused on teaching LC and RC using the TOEIC and TOEFL methods to develop students' comprehensive skills.

TOEIC test

The TOEIC test includes reading and listening, with each test having 100 questions. The reading test takes 75 minutes to complete, including the four sub-tests. Meanwhile, the listening test, which includes three sub-tests, lasts 45 minutes. The TOEIC begins with the listening test and ends with the reading test. The listening test consists of four sub-tests. The first subtest is a listening comprehension test with images. All of the items in this sub-test depict a picture. The test takers are instructed to listen to the statements and select the correct description based on the image.

The second sub-test involves listening to some audio questions and answering multiple-choice questions. The test-takers are instructed to pay close attention to and correctly answer the question. In the meantime, the third and fourth sections involve listening to short conversations to answer a set of questions. The third sub-test requires the test-takers to listen to a brief conversation and read the question. The test-takers then read the possible answers in multiple-choice format and select the correct one. Listening to the short talks is the fourth sub-test. This section asks test takers to listen to several short talks and answer several questions. A report, discussion, announcement, or other functional talks can be used as the test's short talk.

Reading comprehension is another sub-test. Reading comprehension is tested from the fifth to the seventh subtest. The fifth sub-test is incomplete sentences, which require test-takers to select the correct word or phrase. In the sixth sub-test, test-takers must recognize errors, such as identifying incorrect words. The seventh subtest measures reading comprehension. In the seventh section, individuals taking the test engage with diverse reading materials, including letters, advertisements, and notices. They are required to choose the appropriate answers according to the information explicitly or implicitly conveyed in the text.

Test Impact

Several studies have been conducted regarding the TOEIC test impacting the teaching and learning process. The impact of this test is usually called washback (Nguyen & Gu, 2020; Apichatrojanakul, 2011; Wall & Alderson, 1993; Ali & Hamid, 2020; Hung, 2012; Shih, 2010). The term "washback" can also be used in a broader structure, including "the impact of the test on individuals, policies, or practices in the classroom, school, education system, or society (Cheng et al., 2015). In this article, impact assessment refers to improvements in the instructional process brought on by testing results.

(Kuang, 2020) found that one of the main findings of this research is that the analysis of the impact of English language tests indeed requires careful interpretation based on the context. Positive washback from English tests works in line with course syllabuses and classroom teaching to further enhance students' language development while also improving their preparation for exams. Meanwhile, a study by (Indrawati, 2018) concluded that exam preparation learning in certain schools in Tuban applies negative washback. In other words, this research supports the assumptions of previous studies that the washback of the national exam refers to negative washback and has a positive impact on both teachers and students. (Endriyati & Anggraeni, 2019) stated that if the teaching process is oriented to help students pass the National Examination, it falls under negative washback. Furthermore, if the teaching process is oriented to achieve the four language skills listed in the competency standards, it is called positive washback. The intensity of washback is also influenced by the quality of the school. Negative washback is usually found in lower schools and vice versa.

Apichatrojanakul (2011) found positive and negative effects in teaching to increase TOEIC test scores at the Sukhothai Business School in Bangkok, Thailand. This study found that TOEIC instructors must take effective, realistic steps, such as balancing the teaching-centered and student-centered approaches, including paired work and community tasks in the teaching plan, demonstrating the advantages of achieving a high TOEIC score or providing positive feedback to students. Furthermore, successful and acceptable policies and procedures should be established to enhance the positive washback impact and decrease the negative effect on TOEIC teachers at Sukhothai Business School. Teachers should be able to use a communicative style in the classroom by encouraging them to do pair or group work with some listening and reading components that provide the requisite grammar points, making the class more engaging and enjoyable. In contrast, Nguyen & Gu (2020) conducted a study involving non-English students taking the TOEIC Listening and Reading exam as a language exit test; the study indicated that this test could lead to a shift in communicative language teaching and a narrowing of teaching content.

Several studies claim that language tests negatively affect, and modifying the exam will result in positive improvements in language instruction. However, Wall & Alderson (1993) have questioned this claim and have shown that very little empirical research has been conducted to establish the effect of language tests on language teaching. Wall & Alderson (1993) found that positive and negative washback depend on the teaching and teacher content. Similarly, according to the research by (Nahdia, 2017) the results of the language examination implementation have both positive and negative washback effects. The washback can be said to be positive if it is seen as follows. First is Teaching content. Teachers will 'teach textbooks', as they will realize that one type of text or assignment in it may appear on the final exam. Since the weighting of the test will match the textbook's weighting, they will not put more priority on one skill than the

textbook offers. The second is teaching approaches. Teachers will follow the teacher guide's general methodology and strategies because they will see it as an effective means to learn the test's expertise. The third is how to make a decision. The teacher will design a questionnaire that reflects the textbook contents, as this will be the material covered on the exam. They will rate their students' work according to the textbook's standards, which examiners will use when grading O-level tests.

In contrast, a washback can be said to be positive if it is seen as follows: First, the Content of teaching. Teachers might now no longer train the whole textbook because they might recognize that a few talents, particularly listening and speaking, have been no longer assessed and that it turned into extra beneficial to spend constrained magnificence time training analyzing and writing. Even while coaching those talents, instructors may overlook a few textual content kinds or activities, feeling that those in no way regarded on the examination and have been consequently now no longer well worth spending time on. Teachers may abandon using the textbook altogether and start applying different substances that have been extra glaringly associated with the examination. These may encompass teacher-designed substances, beyond exam papers, or guides designed to assist college students in putting together for the examination, and may bring about the narrowing of the curriculum to fit the examination (Smith, [1991a](#)). The second is methodology. Teachers could use any technique they thought would be more helpful in preparing their students for the test. Any aspects of the current textbook's methodology will likely be sacrificed if lecturers believe they are no longer effective in preparing university students. The third is to assess the college students. Teachers will write checks that mimic past exam papers' content rather than the textbook's content. They can change or 'lift' questions from other outlets, such as past papers or study guides, to train college students for the exam. Teachers can focus on the examination's marking criteria and overlook advice from the textbook that contradicts this grading method (Wall & Alderson, 1993). Therefore, teachers must pay more attention to the elements of teaching in the classroom to foster the positive impacts of the test.

Teaching material is more likely to change than the teaching method (Cheng, [1997](#)). Cheng ([1997](#)) mentioned that new materials were written and quickly accepted after introducing the updated Hong Kong Education Examination Certificate (HKCEE). To match the new HKCEE style, the instructor introduced role-plays, group conversations, and more group learning throughout the course. However, the way they interact in the classroom is controlled and dominated by the teacher.

Different assessments can have various impacts on instruction in the identical testing setting. Shohamy et al. ([1996](#)) examined the effects of English as a foreign language (EFL) and Arabic as a second language (ASL) on instruction and exam planning. Shohamy et al. ([1996](#)) found that the impact of the EFL test has risen dramatically, but "the influences of the ASL test has lessened, which indicated the impact of this test was small" (p. 312). In contrast, the identical test provision may have different influences in the distinct school environment. McCartney ([2008](#)) stated practitioners who received high salaries and education levels are free to develop curricula and teaching styles that they find important, allowing them to have a unique teaching effect. Therefore, the research results on the impact of examinations show that introducing a new exam could lead to changes in the content and teaching methods. Consequently, test design must be consistent with relevant abilities or goals, underlying values, teachers' proper interpretation of test objectives and requirements, and instructor preparation and knowledge of acceptable teaching practices (Wall & Alderson, 1993)

The previous research has discussed the impact of TOEIC listening and reading to test scores. It was revealed that TOEIC positively and negatively affected students' test scores (Apichatrojanakul, 2011). It suggested the teacher apply effective and realistic steps to increase students' TOEIC scores. However, previous studies were still limited to the effect of TOEIC on TOEIC scores. It was necessary to research the impact of the TOEIC test on language teaching. Therefore, this research aimed to examine the effect of TOEIC listening and reading tests on English teaching in Indonesian universities and investigate how washback occurs in Indonesia.

Research Context

Indonesia is one of the most linguistically diverse developing countries in the world, with more than 700 languages. With globalization and diversity in the Indonesian economy and increasing international trade, English communication skills are increasingly important, especially for the growing number of Indonesian multinational companies. To improve the quality of the workforce, the government held TOEFL and TOEFL tests for university exit tests. Therefore, this research focuses on the TOEIC test's effect on language teaching by discovering the following questions.

- (1) What is the effect of TOEIC as a graduation requirement on English teaching in Indonesian colleges on language teaching?
- (2) What are teachers' perspectives about factors potentially correlated with the effect of TOEIC as a graduation requirement on English teaching in Indonesian Universities?
- (3) To what extent are factors correlated with the test related to the test impact?

Research Method

This study was a mixed method, consisting of a quantitative and qualitative data. This research aimed to discover the general pattern of the perceived effect of TOEIC listening and reading on teaching TOEIC tests. Therefore, the questionnaire was utilized to examine teachers' views regarding the impact of the test and its related factor. Then, the interview was employed to further examine the survey result.

The questionnaires were distributed to the participants after completing the TOEIC Test on December 10th, 2021. They were asked to respond to the question based on their own experiences and feelings. The administration of the questionnaires took between 5 and 10 minutes. After they had completed the questionnaire, some students were interviewed about the TOEIC Test. Data were collected using a questionnaire and interviews. The interview data were analyzed and compared to the interview results.

Research Side

The research was conducted at three universities in Riau, University A, B, and C. University A is a public institution, while universities B and C are private. University A provided social, medical, and engineering courses, university B had education and social course, and University C offered technology courses. The three universities had dissimilar English learning programs. University A owned three general English modules that adopted communicative learning and integrated skill. The handbook was also completed with TOEIC exercises. The TOEIC benchmark was 450 for all majors. In University B, students learned two general English modules and one TOEIC preparation module. Moreover, they also learned English for Specific Purpose (ESP) in their major, and the requirement to pass was a TOEIC score of 450. In University C,

three English courses are provided and follow the obligation to learn ESP. By the time of data collection for this study, university A has used TOEIC for five years, and universities B and C had three years of experience using it.

Participants

The participants were the teachers who taught TOEIC Listening and Reading course in Universities A, B, and C. The questionnaire was administered to 80 teachers, with a participation rate of 90 %. Thirty-seven were from University A, twenty-eight from university B and 15 students from university C.

Instrument

The questionnaire was developed to discover the impact of TOEIC listening and reading on students' proficiency and the possible factor correlated with the test impact. The survey was meticulously thoroughly. The first step was identifying the variable. Test impact on teaching was defined as the change in teaching content, method, and students' learning strategies due to introducing a new test. Test impact influenced teaching when: 1) there is a time used for direct test preparation, 2) there is a change in teaching methodologies, 3) there is a change in students' learning activities, 4) the increasing of instructional time applied to communicative activities, 5) the increasing of time utilized for tested skill, and 6) time utilized for tested non-skill reduced.

Next, questions were developed based on the variable's description of the element to be evaluated. Several items were adapted from Cheng (1997a). The content validity was done by three independent language experts to assess the clarity and reliability of the instrument. The results showed that the questionnaire was valid to employ as a data collection tool with minor revisions on some items. The next step was the piloting by administering the questionnaire to 30 teachers; they gave comments on the format, their comprehension of the items, their reason for choosing a certain response, and the suitability of one question and another. The pilot study revealed that some formats, instructions, and sentences needed further revision.

The latest version of the questionnaire consisted of two parts. The first part was related to demographic information. The second part consisted of sections A and B. Section A had 65 items with seven Likert scales, and section B related to associated factors that may correlate with the test's impact on students' proficiency. Part 2 used two types of scales: a scale of consensus and a scale of improvement over time: (1) strongly disagree (2) disagree (3) slightly disagree (4) moderate (5) somewhat agree (6) agree, and (7) strongly agree were on the scale of agreement. Furthermore, the time scale of transition decreased greatly (1) mildly (2) marginally (3) remained the same (4) slightly increased (5), moderately increased (6), and substantially increased (7).

The consistency of subscales was assessed by Cronbach's Alpha. Items that decreased the alpha value were deleted until a satisfactory alpha was reached. The reliability test showed the Cronbach's alpha of above 0.70 for most sub-scales, except for learning activity (0.65), test difficulty (0.68), belief in the necessity of test preparation (0.57), context factor-opportunities (0.54).

Data collection procedure

The Questionnaire was distributed during class hours. The teacher briefed the participants about the questionnaire and requested their willingness to participate in this research. To encourage honesty in filling out the questionnaire, the teacher was not present in the class.

Analysis

The data were analyzed using SPSS to examine the durability and item's accuracy. The cumulative score for each subscale was obtained by adding all item scores. The overall score was then divided by the number of objects in each subscale to translate the score to seven scales. Next, the questionnaire was analyzed using descriptive statistics (mean and standard deviation); the testing effect and related components may be affected according to the mean. A bivariate correlation analysis was also used to assess the test-correlated factor association.

Result and Discussion

Perceive the impact of TOEIC Listening and Reading

Table 1 shows that students tend to agree that TOEIC learning and reading impacted direct test preparation (M=5,58, SD=0,66). It was found that the test changed the teaching methodology (M = 5.81, SD = 0.69). It was also known that the teacher encouraged students to change their learning strategies (M = 5.39, SD = 0.71). Besides, time devoted to communicative activities (M = 5.41, SD = 0.71), tested skill (M = 5.43, SD = 0.87), and non-tested skill (M = 5.82, SD = 0.79) were also increased. The test's effect on time spent on non-tested skills and changes in teaching methods was the greatest, close to 6. Since the standard deviations were minimal, these results were comparable for most teachers (0.69 and 0.79, respectively). Teachers' perceptions of time spent on communicative tasks and tested abilities, on the other hand, differed widely, as shown by the highest standard deviations (0.90 and 0.87, respectively).

Table 1. Perceived impact of TOEIC L&R on learning

Test impact	N	Min	Max	Mean	SD
Direct test preparation	80	3.50	6.75	5.58	0.66
Change of teaching methodology	80	4.00	7.00	5.81	0.69
Change of students' learning strategies	80	3.50	6.75	5.39	0.71
Time utilized for communicative activities	80	3.33	7.00	5.41	0.90
Time utilized for tested skill	80	3.20	6.60	5.43	0.87
Time utilized for non-tested skill	80	3.67	7.00	5.82	0.79

The teacher's responses in the interview session confirmed the inclination to provide instruction specifically tailored for the test. They stated, "I only highlighted tested the skills and knowledge necessary for students to excel in the TOEIC test so that students were able to have a big chance to practice" (P5). They also added, "the increasing use of strategies to assist the student to select the correct answer" (P18). Regarding the influences of TOEIC listening and reading on the utilization of teaching methods, teachers explained they were becoming less communicative during teaching practice. "TOEIC impacted the decreasing time to plan communicative tasks in class," admitted the teacher (P11). The utilization of the English language in a class has decreased, according to the teacher (P9): "The use of English for studying and teaching in TOEIC class has decreased because the exam format does not include speaking." As a result, "Since the teaching and learning process only included students filling out multiple-choice items, the course was less creative" (P3).

On the other hand, TOEIC positively impacts students' language ability in English (Woodford, 1980), and TOEIC listening and reading score strongly correlate with students' receptive vocabulary. It confirms that vocabulary size affects the TOEIC listening and reading score (Chiang, 2018). Furthermore, students agreed that using the TOEIC test score as a graduation prerequisite benefits their language proficiency and

future employment prospects. Studying for the test positively impacts their language proficiency and future employment prospects (Hsieh, 2017; Chen & Hung, 2014). Therefore, students need to prepare to take a test and motivate themselves to utilize listening strategies to get satisfactory TOEIC scores (Lee, 1990).

Factor Associated with Perceived Impact of TOEIC Listening and Reading

Table 2 revealed teachers' beliefs on the communicative approach (M=5,55, SD=0,92) and traditional approach (M=5,85, SD=0,79). Besides, participants had the same opinion about the importance of direct test preparation (M = 5.30, SD = 0.92) to enhance test scores. The findings also confirmed teachers recognized the purpose of the test and the test format well. They believed the test would evaluate the tested skills (M = 5.61, SD = 0.99) and not assess non-tested skills (M = 5,63, SD = 0.94). They were convinced with the test design (M = 5.70, SD = 0.89), comprehended the test significance (M = 5.74, SD = 0.58), and were aware of the complexity of the test (M = 5.70, SD = 0.98).

Table 2. Factors Related to the Perceived of the Impact of TOEIC L&R on Teaching

Test impact	N	Min	Max	Mean	SD
(1) Principles of teaching and learning language	80	3.25	6.75	5.30	0.92
The idea of a communicative approach	80	3.00	7.00	5.55	0.91
The idea of conservative methods	80	3.50	6.50	5.85	0.79
Belief in direct test preparation to improve test scores	80	3.25	6.75	5.30	0.92
(2) Response on TOEIC L&R					
Test purpose–tested skills	80	3.43	6.71	5.61	0.99
Test purpose–non-tested skills	80	3.33	7.00	5.65	0.94
Test design	80	3.67	7.00	5.70	0.89
Test significance	80	4.00	7.00	5.74	0.58
Test complexity	80	3.33	7.00	5.70	0.98
Consequences for teachers	80	3.25	6.50	4.78	1.01
External expectations	80	3.25	7.00	4.68	1.06
(3) Response to context factor					
Limitation	80	3.50	6.25	5.07	0.96
Chances	80	3.40	6.00	4.54	0.98

It was found that teachers had comprehended the design of the test, but they found that it had a positive effect on the learning strategies. For instance, one teacher stated: I have realized TOEIC could encourage students to apply effective and efficient learning strategies (P15). It corroborated teaching agreement on the significance of the test (Table 2). The results of the questionnaire on the test significance remarked on the teacher's belief in the significance of the TOEIC test (M = 5.74, SD = 0,58). Teachers then assumed TOEIC L&R was just marginally difficult (M = 5.70, SD = 0.98). They had no prior research pressure experience. They thought the test had a minor impact on them (M = 4.78, SD = 1.01) and that external goals were only marginally met (M = 4.68, SD = 1.06). Furthermore, they earned more demands from stakeholder expectations (M = 5.07, SD = 0.96) than direct effects on themselves (M = 4.54, SD = 0.98). (Table 2). Both limitations and possibilities existed for the instructor (Table 2). The limitations included a class size and time constraints, prohibiting them from teaching particular material or planning specific classroom events. Meanwhile, opportunities were teachers' freedom to utilize textbooks and access varied learning sources.

Test Factor

Important associations existed between test factors and the perceived test effect (Table 3). Perception of test value, test purposes, and external expectation were all factors linked to test effects. Teachers' perceptions of test architecture are compatible with practices. For instance, test format ($r=0.19$) was significantly associated with direct test preparation. The test complexity negatively correlated with time utilized for communicative activities ($r= -0.19$), indicating that the more difficult the test, the less time used for communicative activities. In addition, the external expectation was negatively associated with the change in teaching methodology ($r= -0.21$).

Table 3. Correlations between Test Factors and Perceived Impact of TOEIC L&R

		Test significance	Test purpose tested skills	Test Purpose Non- Tested Skills	Test Design	Test complexity	Consequence for Teacher	External Expectation
Direct Test Preparation	Pearson Correlation	-0.038	-0.072	-0.062	0.199*	-0.012	0.149	0.065
	Sig. (2-tailed)	0.740	0.525	0.582	0.077	0.919	0.188	0.569
	N	80	80	80	80	80	80	80
Change of teaching methodology	Pearson Correlation	-0.005	0.004	-0.029	-0.158	-0.024	-0.119	-0.212*
	Sig. (2-tailed)	0.964	0.974	0.795	0.161	0.834	0.294	0.059
	N	80	80	80	80	80	80	80
Change of students' learning strategies	Pearson Correlation	-0.110	-0.071	-0.128	0.175	0.007	0.007	-0.110
	Sig. (2-tailed)	0.332	0.533	0.257	0.120	0.949	0.952	0.333
	N	80	80	80	80	80	80	80
Time utilized for communicative activities	Pearson Correlation	0.007	0.045	0.009	0.011	-0.191*	-0.055	-0.071
	Sig. (2-tailed)	0.948	0.692	0.936	0.924	0.090	0.629	0.529
	N	80	80	80	80	80	80	80
Time utilized for tested skill	Pearson Correlation	-0.074	-0.075	-0.057	-0.010	0.093	0.038	0.039
	Sig. (2-tailed)	0.511	0.506	0.613	0.932	0.411	0.735	0.731
	N	80	80	80	80	80	80	80
Time utilized for non-tested skill	Pearson Correlation	0.085	-0.045	-0.095	0.185	0.069	0.152	0.149
	Sig. (2-tailed)	0.452	0.693	0.403	0.101	0.543	0.178	0.187
	N	80	80	80	80	80	80	80

*Correlation is significant at the 0.1 level (2-tailed).

**Correlation is significant at the 0.05 level (2-tailed).

Table 3 reveals that test preparation had the biggest influence on students' perceived impact. Test preparation has proved to affect test scores to varying grades. It seems that there was no significant difference in the approach to preparing tests (O'Sullivan et al., 2019), and it is clear that mastery goals are related to test preparation (Razavipour et al., 2020). When it comes to testing planning, teachers have various options, including a) a standard curriculum with no preparation and a customized curriculum with preparation, b) teaching test-taking skills, c) making suggestions, d) teaching test material, e) teaching test structure and content, f) stress vaccination, g) test preparation or parallel test pieces, and h) cheating (Smith, 1991b). Even though test preparation improves test scores, it leads to the narrowing of the curriculum, especially drilling (Xie, 2013).

Besides, test preparation should deal with the issue of program and policy. The program of test preparation should include the preparation of students' vocabulary because it will affect their vocabulary knowledge and confidence in taking the test (Ching-Shyang Chang, 2007); it is also important to provide test preparation according to students' level, such as instruction for higher level (Park et al., 2020). Test influences language policies, leading to the view of language as standardized and homogenous. However, a test can be a tool to create valid language policies that mediate the ideology and practice (Shohamy, 2007). Then, teachers are better prepared to encounter the tension between the critical view of the high-stake test adopted by TEFL and the critical view of institutional policies (Johnson et al., 2005).

Person Factor

Table 4 reveals that age was significantly correlated with time utilized for tested skills ($r= 0.24$) but negatively associated with the change in students' learning strategies ($r= -0.21$). In addition, experiences in teaching were correlated with time utilized for tested skills ($r=0.20$) but negatively linked to direct test preparation ($r= -0,20$). Age was not correlated to the change of teaching methodology, time utilized for communicative activities, and time utilized for non-tested skills. Teaching experiences also were not associated with changes in teaching methodology, changes in students' learning strategy, and time devoted to tested skills.

Experience in TOEIC preparation positively correlated with the time utilized for communicative activities ($r= 0,190$) and tested skills ($r= 0,195$). This suggested that teachers with more TOEIC training expertise would devote more time to communicative tasks and less time to the tested ability. Surprisingly, the testing effect was unrelated to assumptions about language instruction. Test impact had no correlation with the communicative approach or conservative method.

Table 4. Correlations between Person Factors and Perceived Impact of TOEIC L&R

		Age	Experience in teaching	Experience in TOEIC	The idea of Communicative approach	Idea of Conservative Methods	Direct Test belief
Direct Test Preparation	Pearson Correlation	-0.091	-0.202*	-0.116	0.034	-0.074	0.049
	Sig. (2-tailed)	0.420	0.072	0.307	0.761	0.513	0.664
	N	80	80	80	80	80	80

Change of teaching methodology	Pearson Correlation	0.184	-0.044	0.117	0.123	0.144	-0.049
	Sig. (2-tailed)	0.103	0.699	0.303	0.279	0.202	0.668
	N	80	80	80	80	80	80
Change of students' learning strategies	Pearson Correlation	-0.217*	-0.082	-0.122	-0.082	0.058	0.008
	Sig. (2-tailed)	0.053	0.467	0.281	0.470	0.608	0.944
	N	80	80	80	80	80	80
Time utilized for communicative activities	Pearson Correlation	-0.076	0.047	0.190*	0.067	-0.126	0.112
	Sig. (2-tailed)	0.504	0.679	0.091	0.557	0.266	0.324
	N	80	80	80	80	80	80
Time utilized for tested skill	Pearson Correlation	0.245**	0.200*	0.195*	0.153	-0.042	0.041
	Sig. (2-tailed)	0.029	0.075	0.083	0.175	0.710	0.721
	N	80	80	80	80	80	80
Time utilized for non-tested skill	Pearson Correlation	0.086	-0.095	0.116	0.088	-0.103	-0.327**
	Sig. (2-tailed)	0.450	0.400	0.304	0.436	0.361	0.003
	N	80	80	80	80	80	80

*Correlation is significant at the 0.1 level (2-tailed).

**Correlation is significant at the 0.05 level (2-tailed).

Moreover, test takers performance on the TOEIC test was positively correlated with their ability to carry out English language tasks encountered in the workplace and everyday life and has been recognized as a framework describing language proficiency levels (Powers, 2018). Besides, teachers also have an important role related to the perceived impact. The teacher can discover how knowledge of the test construct can tell their choices related to the curriculum, material, teaching methods, feelings, attitudes and students' learning to create a positive test impact (Schissel, 2018).

Furthermore, test takers' experience is also correlated with perceived impact. Most students did not have enough experience with the standardized test, especially TOEIC, and did not prepare for the test. They had problems with the listening section due to the limited vocabulary, and they faced problems concerning the meaning of the context of the text (Zahruni et al., 2020). The students who achieve a low score on the test become anxious about the test, cheat or use poor test-taking strategies. These reactions may defend the feeling of competence when they accept a low score (Paris et al., 1991). Additionally, some students are under pressure to take the EFL test, and secondary institutions emphasize the preparation for the exam, resulting in students' negative views of the test (Choi, 2008). Other research found teachers and students feel

the pressure related to high-stake testing, which is the effect of the drill and practice type of curriculum. In addition, promising students were under pressure to perform well and raise their overall grades, resulting in disengagement from the learning process (Moon et al. 2007).

Context Factor

Table 5. Correlations between Context Factors and Perceived Impact of TOEIC L&R

		Constraints	Chance
Direct Test Preparation	Pearson Correlation	0.082	-0.315**
	Sig. (2-tailed)	0.470	0.004
	N	80	80
Change of teaching methodology	Pearson Correlation	0.079	-0.032
	Sig. (2-tailed)	0.487	0.777
	N	80	80
Change of students' learning strategy	Pearson Correlation	0.061	-0.006
	Sig. (2-tailed)	0.591	0.955
	N	80	80
Time utilized for communicative activities	Pearson Correlation	0.044	-0.135
	Sig. (2-tailed)	0.697	0.232
	N	80	80
Time utilized for non-tested skill	Pearson Correlation	-0.007	0.000
	Sig. (2-tailed)	0.948	0.997
	N	80	80
Time utilized to tested-skills	Pearson Correlation	0.155	0.022
	Sig. (2-tailed)	0.171	0.849
	N	80	80

*Correlation is significant at the 0.1 level (2-tailed).

**Correlation is significant at the 0.05 level (2-tailed).

Using TOEIC as an exit test requirement encourages the students to learn English instead of impacting the teaching. Data from a larger study indicated that TOEIC as a graduation requirement increased students' motivation to learn English. However, no consistent policy regarding curriculum narrowing or a communicative teaching style existed. Using TOEIC speaking and writing or other authentic teaching and learning activities will be helpful. In other words, an exit exam should be linked to a curricular target such that teaching and learning promote progress toward the goal. Furthermore, the exit requirement test is structured to ensure students' English proficiency. It is based on various criteria, such as face authenticity, test-taker characteristics, feasibility, and local contextual problems. Many of these considerations were restricting because the four skills tested on the TOEIC were unaffordable and unattainable for most students, and there was no locally accessible produced content.

Conclusion

This study revealed the power of the test in influencing teaching content and teaching method. Furthermore, this analysis showed that the teaching for the test did not involve communicative teaching because it did not comply with the program objectives.

However, the students should use English communicatively. A potential alternative is to administer the entire version of TOEIC to the pupils, including tests of productive skills, such as speech and writing. There are some drawbacks to this report. First, since the new results focus on teachers' perspectives, they are transient. Second, two subscales had just two products, and Cronbach's alphas of some were below 0.7. The thesis then concentrated solely on the effect of the exam on teaching rather than studying. Finally, since the participants were only from three universities in Indonesia, the finding could not be applied to other universities in Indonesia. Further research should ask some questions to fully comprehend the testing effect on the Indonesian context, for example, examining the connection between age, teaching experience, and test effect.

References

- Ali, M. M., & Hamid, M. O. (2020). Teaching English to the Test: Why Does Negative Washback Exist within Secondary Education in Bangladesh? *Language Assessment Quarterly*, 17(2), 129–146. <https://doi.org/10.1080/15434303.2020.1717495>
- Apichatrojanakul, P. (2011). The Washback Effects of the TOEIC Examination on the Teachers and Students of a Thai Business School. *Language Testing in Asia*, 1(1), 62–75. <https://doi.org/10.1186/2229-0443-1-1-62>
- Booth, D. K. (2018). *The TOEIC* (pp. 19–35). Springer, Cham. https://doi.org/10.1007/978-3-319-70446-3_3
- Chen, M.-L., & Hung, L.-M. (2014). *Influence of TOEIC on Teachers and Students at Tertiary Level*. 93–94. <https://doi.org/10.2991/icasr-14.2014.26>
- Cheng, L. (1997). How does washback influence teaching? Implications for hong kong1. *Language and Education*, 11(1), 38–54. <https://doi.org/10.1080/09500789708666717>
- Cheng, L., Sun, Y., & Ma, J. (2015). Review of washback research literature within Kane's argument-based validation framework. In *Language Teaching* (Vol. 48, Issue 4, pp. 436–470). Cambridge University Press. <https://doi.org/10.1017/S0261444815000233>
- Chiang, H.-H. (2018). English Vocabulary Size as a Predictor of TOEIC Listening and Reading Achievement among EFL Students in Taiwan. *Theory and Practice in Language Studies*, 8(2), 203. <https://doi.org/10.17507/tpls.0802.04>
- Ching-Shyang Chang, A. (2007). The impact of vocabulary preparation on L2 listening comprehension, confidence and strategy use. *System*, 35(4), 534–550. <https://doi.org/10.1016/j.system.2007.06.003>
- Choi, I.-C. (2008). The impact of EFL testing on EFL education in Korea. *Language Testing*, 25(1), 39–62. <https://doi.org/10.1177/0265532207083744>
- Endriyati, R., & Anggraeni, D. A. (2019). The Washback Effect of National Examination on English Language Teaching in Junior High School in Indonesia. *The 2nd International Conference on Language, Literature and Teaching*, 106–111.
- Hsieh, C.-N. (2017). The Case of Taiwan: Perceptions of College Students About the Use of the TOEIC® Tests as a Condition of Graduation. *ETS Research Report Series*, 2017(1), 1–12. <https://doi.org/10.1002/ets2.12179>
- Hung, S. T. A. (2012). A washback study on e-portfolio assessment in an English as a Foreign Language teacher preparation program. *Computer Assisted Language Learning*, 25(1), 21–36. <https://doi.org/10.1080/09588221.2010.551756>
- Indrawati, T. (2018). Wash-Back of English National Examination on Teaching

- Learning Process. *RETAIN*, 6(3), 60–68.
- Johnson, K. E., Jordan, S. R., & Poehner, M. E. (2005). The TOEFL Trump Card: An Investigation of Test Impact in an ESL Classroom. *Critical Inquiry in Language Studies*, 2(2), 71–94. https://doi.org/10.1207/s15427595cils0202_1
- Jung, H. (2010). Washback Effects of New TOEIC on Korean College Students' English Teaching and Learning. *English21*, 23(3), 183–207. <https://doi.org/10.35771/engdoi.2010.23.3.009>
- KI Shin, & Oh. (2012). The effects of TOEIC Teaching as a Compulsory Course in a College. *Studies in Linguistics*, null(24), 125–137. <https://doi.org/10.17002/sil..24.201207.125>
- Kuang, Q. (2020). A Review of the Washback of English Language Tests on Classroom Teaching. *English Language Teaching*, 13(9), 10–17. <https://doi.org/10.5539/elt.v13n9p10>
- Lee, J.-A. (1990). 듣기 전략 사용 선호도가 TOEIC 듣기 성취도에 미치는 영향과 매개 변인과의 관계 *)에 의해 학습전략 목록과 분류 체계가 제시되고 측정. In *English Language & Literature Teaching* (Vol. 15, Issue 4). The English Teachers Association in Korea.
- Mahaputri, D. S. (2013). Improving Sophomore Listening Ability By Using Toeic and Toefl Exercises At Stkip Abdi Pendidikan Payakumbuh. *Al-Ta Lim Journal*, 20(3), 524–533. <https://doi.org/10.15548/jt.v20i3.49>
- McCarthy, S. J. (2008). The Impact of No Child Left Behind on Teachers' Writing Instruction. *Written Communication*, 25(4), 462–505. <https://doi.org/10.1177/0741088308322554>
- Moon, Tonya R.; Brighton, Catherine M.; Jarvis, Jane M.; Hall, C. J. (2007). THE NATIONAL RESEARCH CENTER ON THE GIFTED AND TALENTED State Standardized Testing Programs: Their Effects on Teachers and Students. In *National Research Center on the Gifted and Talented*. National Research Center on the Gifted and Talented. University of Connecticut, 2131 Hillside Road, Unit 3007, Storrs, CT 06269-4676. Tel: 860-486-4676; Fax: 860-486-2900; Web site: <http://www.gifted.uconn.edu/nrcgt.html>.
- Nahdia, K. (2017). Washback Analysis of Students' Perception and Teachers' Teaching Material of English National Examination 2017. *Journal of English Language Teaching*, 6(1), 92–101. <https://doi.org/10.15294/elt.v6i1.20685>
- Nguyen, H., & Gu, Y. (2020). Impact of TOEIC Listening and Reading as a University Exit Test in Vietnam Impact of TOEIC Listening and Reading as a University Exit Test in. *Language Assessment Quarterly*, 00(00), 1–21. <https://doi.org/10.1080/15434303.2020.1722672>
- O'Sullivan, B., Dunn, K., & Berry, V. (2019). Test preparation: an international comparison of test takers' preferences. *Assessment in Education: Principles, Policy & Practice*, 1–24. <https://doi.org/10.1080/0969594x.2019.1637820>
- Paris, S. G., Lawton, T. A., Turner, J. C., & Roth, J. L. (1991). A Developmental Perspective on Standardized Achievement Testing. *Educational Researcher*, 20(5), 12–20. <https://doi.org/10.3102/0013189X020005012>
- Park, S., Kwak, E.-J., Tak, J., & Tate, T. (2020). Investigation on TOEIC score trends in Korea and its pedagogical implications. *Cogent Education*, 7(1), 1796557. <https://doi.org/10.1080/2331186X.2020.1796557>
- Powers, D. E. (2018). TOEIC (Test of English for International Communication). In *The TESOL Encyclopedia of English Language Teaching* (pp. 1–8). John Wiley &

- Sons, Inc. <https://doi.org/10.1002/9781118784235.eelt0630>
- Powers, D. E., Kim, H.-J., & Weng, V. Z. (2008). THE REDESIGNED TOEIC® (LISTENING AND READING) TEST: RELATIONS TO TEST-TAKER PERCEPTIONS OF PROFICIENCY IN ENGLISH. *ETS Research Report Series*, 2008(2), i–120. <https://doi.org/10.1002/j.2333-8504.2008.tb02142.x>
- Razavipour, K., Habibollahi, P., & Vahdat, S. (2020). Preparing for the higher education admission test: preparation practices and test takers' achievement goal orientations. *Assessment and Evaluation in Higher Education*. <https://doi.org/10.1080/02602938.2020.1773392>
- Schissel, J. L. (2018). Test Impact and Washback. In *The TESOL Encyclopedia of English Language Teaching* (pp. 1–6). John Wiley & Sons, Inc. <https://doi.org/10.1002/9781118784235.eelt0353>
- Shih, C. M. (2010). The washback of the general English proficiency test on university policies: A Taiwan case study. *Language Assessment Quarterly*, 7(3), 234–254. <https://doi.org/10.1080/15434301003664196>
- Shohamy, E. (2007). Language tests as language policy tools. *Assessment in Education: Principles, Policy and Practice*, 14(1), 117–130. <https://doi.org/10.1080/09695940701272948>
- Shohamy, E., Donitsa-Schmidt, S., & Ferman, I. (1996). Test impact revisited: washback effect over time. *Language Testing*, 13(3), 298–317. <https://doi.org/10.1177/026553229601300305>
- Smith, M. L. (1991a). Put to the Test: The Effects of External Testing on Teachers. *Educational Researcher*, 20(5), 8–11. <https://doi.org/10.3102/0013189X020005008>
- Smith, M. L. (1991b). Meanings of Test Preparation. *American Educational Research Journal*, 28(3), 521–542. <https://doi.org/10.3102/00028312028003521>
- Wall, D., & Alderson, J. C. (1993). Examining washback: The Sri Lankan Impact Study. *Language Testing*, 10(1), 41–69. <https://doi.org/10.1177/026553229301000103>
- Wilson, K. M. (2000). AN EXPLORATORY DIMENSIONALITY ASSESSMENT OF THE TOEIC TEST. *ETS Research Report Series*, 2000(2), i–28. <https://doi.org/10.1002/j.2333-8504.2000.tb01837.x>
- Woodford, P. E. (1980). *The Test of English for International Communication (TOEIC)*.
- Xie, Q. (2013). Does test preparation work? implications for score validity. *Language Assessment Quarterly*, 10(2), 196–218. <https://doi.org/10.1080/15434303.2012.721423>
- Yahmun, Y., Sumarti, E., & Setyowati, D. (2020). LISTENING DIFFICULTIES FACED BY THE FIRST SEMESTER STUDENTS AT BASIC LISTENING CLASS. *Journey: Journal of English Language and Pedagogy*, 3(1), 57–61. <https://doi.org/10.33503/journey.v3i1.710>
- Zahrani, N. A., Fahmi, F., & Pratolo, B. W. (2020). The Challenges of Taking TOEIC Test and How to Overcome: Perception of Indonesian Vocational Students. *Ethical Lingua: Journal of Language Teaching and Literature*, 7(1), 82–91. <https://doi.org/10.30605/25409190.167>