Research article

The washback effect of TOEFL iBT and a local English Proficiency Exam on students' motivation, autonomy and language learning strategies

Karim Sadeghi, Aylin Ballidak, Enisa Medea

A Department of English Language Teaching, Bahçeşehir University, Istanbul, Turkey
B English Language Preparatory School, Istanbul Technical University, Istanbul, Turkey

ARTICLE INFO

Keywords:
Washback
High-stakes tests
TOEFL-iBT
English proficiency exam
Motivation
Autonomy, language learning strategies
Language testing

ABSTRACT

This study aimed to investigate the washback effect of two high stakes tests, a global language proficiency test (i.e., TOEFL iBT) and a local English Proficiency Exam (developed and administered by a state university) on students' motivation and their autonomy. The study also examined whether proficiency level moderated the potential washback effect among the two groups of test takers. Additionally, the study tried to find out the language leaning strategies used by both groups and explore the reasons behind their preferences. The study was conducted with two English language preparatory programs offered at a state university in Turkey: University Preparatory Program (UPP) and Dual Degree Program (DDP). At the end of the UPP program, the students are required to take the university's proficiency test while as for the UPP, they need to take a valid TOEFL iBT. Data were collected from 152 preparatory students (N = 65 for DDP; N = 87 for UPP) whose proficiency levels were based on the CEFR Framework ranging from A2 (upper elementary) to B1 (pre-intermediate) to B2 (intermediate). In addition to the above proficiency tests, data were gathered via motivation questionnaire, autonomous learning scale and student interviews. The results revealed no washback of TOEFL iBT exam on students' motivation regardless of their proficiency level. No washback was also observed on students' autonomy except for A2 level DDP students who had higher level of autonomy than the A2 level UPP students. Finally, the two groups used more similar than different language learning strategies while practicing the four language skills. The findings afford pedagogical implications for the use of high-stakes tests in English preparatory programs.

1. Introduction

Having a long history, testing dates back to the Chinese Imperial examinations as it was the first state-wide attempt for a centrally controlled exam (Spolsky, 2017). In the education context, tests have a crucial role because of their potential impact on learners' achievement when exposed to specific teaching programs. Whether testing is done with or without explicit purposes, its impact on teaching/learning and learners is undeniable. Therefore, various changes could be observed in teaching and learning processes and products regardless of the primary aims of test designers because of the complicated relationship between the different forces inside and outside of the school (Cheng and Curtis, 2004).

The washback hypothesis, proposed by Alderson and Wall (1992), assumes that “teachers and learners do things they would not necessarily otherwise do because of the test” (p. 6). The researchers argued that washback could not be considered the same with the influence of tests on teaching and learning because what happens in the classroom are not the things only caused by the test itself. They made distinction between test impact and washback, with the former being a broader area, which constitutes a crucial place to figure out what actually happens in the classroom. To be more precise, washback was thought to be one aspect of impact by language examiners (Tsagari, 2007). Recently, Alderson (2004) stated that washback was such a complex issue that it was a lot more than the negative consequences of tests; however, there was no hesitation about the existence of washback. Similarly, Watanabe (2004) emphasised the dimensions of washback as, specificity, intensity, length, intentionality, and value indicated that washback could be caused by any test or type of test or a specific characteristic of a test.

Given the fact that there are many issues related to washback, it is not surprising that each test's impact will be different from that of the others. The main feature that may change the outcomes of a test is its stakes. High-stakes tests are considered to have more substantial effects compared to those with lower stakes (Menken et al., 2017). Although the

https://doi.org/10.1016/j.heliyon.2021.e08135
Received 10 May 2021; Received in revised form 21 August 2021; Accepted 1 October 2021
2405-8440/© 2021 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
nature of tests cannot be identified easily since the consequences may or may not be the same for different people (Alderson, 2004). As high-stakes tests have more visible results and since they need a higher effort to succeed in, they can lead to modifications in educational behaviours and plans (Shohamy, 2017).

2. Theoretical background of washback

Washback, also referred as backwash, was simply described as the influence of tests by Hughes (1989, p. 1), and it has taken attention of many scholars since then (Alderson and Wall, 1993; Pan, 2014; Watanabe, 2004). Alderson and Wall (1993), who undertook one of the preliminary studies, analysed the term washback and pointed to a difference between test impact and washback by stating that impact is a term that includes many other dynamics. The impact of a test is described as a test’s effect on people, regulations or techniques in different contexts varying from classrooms to the whole society, and today washback is considered to be one aspect of impact (Tsagari and Cheng, 2017).

Alderson and Wall (1993) proposed 14 hypotheses of washback. A test will influence: (a) teaching, (b) learning, (c) what teachers teach, (d) how teachers teach, (e) what learners learn, (f) how learners learn, (g) the rate and sequence of teaching, (h) the rate and sequence of learning, (i) the degree and depth of teaching, (k) degree and depth of learning, (l) attitudes to content, method, etc. of teaching/learning. Also, (m) tests that have important consequences will have washback, (n) tests that do not have important consequences will have no washback, (o) tests will have washback on all learners and teachers, and (p) tests will have washback effects for some teachers and some learners, but not for others (p. 120).

Besides, Hughes (1993) suggested a trichotomy of participants, process and product as basic constructs of backwash. Based on the proposed model, the perceptions and attitudes of participants towards their teaching and learning tasks might be affected primarily by the nature of a test. As a result, the process that participants undergo might influence the learning outcomes as well as the product.

To summarize, it can be inferred that washback has the potential for affecting not only the learner but also the program and the education system. Hence, the impact of tests on the process of teaching and learning should be closely addressed to gain a deeper understanding of the nature of washback and how it varies from test to test and from context to context.

2.1. The nature of washback

The notion of “washback” is a vital concept in the field of language teaching and assessment. According to Alderson and Wall (1993), washback occurs when teachers and learners do things which they would not necessarily if there was no test. Similarly, Bailey (1996) argues that washback is the influence of testing on the process of teaching and learning. Pearson (1998) also emphasizes the fact that public examinations are influential on the attitudes, behaviors, and motivation of teachers, learners, and also parents. As examinations are often given at the end of a course, this influence has a backward direction, hence washback. Last but not least, washback is related to an intended or unintended (accidental) change and functions parallel to the curriculum change in the educational context (Cheng, 2005).

Although language tests have often been considered to have a negative influence on teaching, washback could be both negative and positive (Alderson and Wall, 1993; Bailey, 2005; Hughes, 1989; Taylor, 2005). Negative washback is the possible unwanted outcome of a specific test on teaching and learning (Alderson and Wall, 1993) while “positive washback encourages good teaching practices” following the test (Taylor, 2005, p. 154). The common misconception is that the type of washback is directly related to the quality of the test. However, as opposed to this overgeneralization, the relationship between positive and negative washback cannot be claimed to be linear with good and bad tests. Alderson and Wall (1993) stated that as long as the test causes desirable effects, it could have beneficial washback no matter it is a good or bad test; for instance, the teachers and the learners may take some actions like giving more importance to the topic, better lesson preparation or doing homework as a result of a poor test, which may be considered as beneficial, therefore as a positive washback. Thus, positive and negative washback are bidirectional and affected by various contextual factors, so when the washback of a specific test is investigated, a good knowledge of the educational context is necessary (Cheng and Curtis, 2004).

2.2. Washback and motivation

With the increasing research undertaken on washback, the relationship between affective factors such as attitudes, motivation, self-confidence and achievement has been addressed by distinguished scholars (Alderson and Wall, 1993; Booth, 2018; McKenzie, 2008). Although motivation is generally considered as one crucial determinant of language performance, the motivational factors that are influential in a test-oriented learning environment and whether they have an impact on test performance are still under investigated (Wu and Lee, 2017).

Motivation research address the reasons that lead a person to make certain decisions, take action and be consistent with it (Dörnyei and Ushioda, 2013). As categorized by Eccles and Wigfield (2002), there are theories focusing on reasons for task engagement (e.g., intrinsic motivation theory, self-determination theory), theories on integrating expectancy and value constructs (e.g., attribution theory, expectancy-value theory) as well as theories that link motivation with cognition (e.g., theories of self-regulation and motivation). In line with these theories, Dörnyei (2005) divided research on motivation into certain periods: the social psychological period, the cognitive-situated period and the process-oriented period. The initial studies were done by social psychologists like Gardner (1985) who emphasized three components of the integrative motivation model namely, integrativeness, attitudes toward the learning situation and motivation. According to Gardner (1985), individuals have a motivated organism only when the desire to attain the goal and favorable attitudes are associated with the effort or the drive. Additionally, the focus of motivation research moved towards relationship with contextual factors, which paved the way for Self-Determination Theory (SDT) (Deci and Ryan, 1985) highlighting the importance on different types of motivation. Following that, the process-oriented period, as the name suggests, focused on the change in the motivation in time. Dörnyei and Otto (1998) developed a motivation model in which they carefully analyzed different stages of motivation.

Last but not least, washback might be positively influential on teaching and learning practices. As stated, positive washback may be fostered by changing the factors of test design content and interpretation and also, adjusting factors of test logistics. Similarly, Brown (2004) approached positive washback as “the information that ‘washes back’ students diagnosing their strengths and weaknesses and also, enhancing factors such as intrinsic motivation, autonomy, self-confidence, inter-language, and strategic investment” (p. 29). In this sense, good tests will positively influence the teaching-learning process, enhance motivation among teachers and learners as well as encouraging the idea of lifelong learning (Pan, 2009). In brief, as washback may have both positive and negative effects, it is important to investigate the effects of high-stakes tests on learner language motivation in general.

2.3. Washback and learner autonomy

Learner autonomy is defined as learner independence, self-direction, autonomous learning and independent learning (Palfreyman, 2005, p. 3). Various factors play a role in shaping a person’s autonomy such as, individual differences like needs, interests, learning habits or motivation which provide students with freedom in their learning experience (Bildi, 2017). Factors like motivation should be taken into consideration
because self-motivation is influential for autonomous learning. When learners start finding the grounds for their language learning and specify their objectives, it will raise their level of self-motivation. Given that motivation and autonomy are closely connected, and the standing of autonomous learning in proficiency development, autonomy was another variable explored in this study.

2.4. Washback and language learning strategies

Strategy was referred by Oxford and Crookall (1989) as “learning techniques, behaviors, or actions; or learning to learn, problem solving or study skills” (p. 404). Similarly, strategy has been described as “a conscious technique used by a learner purposely to assist the language learning process” (Grainer, 2005, p. 328). Analyzing different descriptions of this term, Purpura (1999) highlighted four concerns in defining strategy: the precise nature of the behaviors that makes a strategy, the issue of observability, the issue of intentionality, and the directness of the learning or performance.

Early focus of strategy use has emphasized its importance in the process of learning. According to Woolfolk (1998), learning strategy is “a plan implemented to achieve the learning objectives” (p. 78). Similarly, learning strategies have been defined as “the processes which are consciously selected by learners and which may result in action taken to enhance the learning or use of a second or foreign language” (Cohen, 1998, p. 4) which facilitates the ‘process of self-learning for the student’ (Ozer, 2002, p. 17). In short, the better the language learners are, the more strategic they are compared to their fewer effective counterparts (Chamot, 1998). Therefore, when learners are equipped with the necessary strategies, they have greater potential to achieve success in language tests.

Cohen and Weaver (2006) classified the use of strategies depending on the language skills favored by the learner such as productive (writing and speaking) or receptive (reading and listening). Nonetheless, general skill-related strategies as vocabulary, grammar, or translation strategies are also employed across different skills. In addition, the use learning strategies can be dependent on the learner’s goal as well.

Considering the stages in test preparation, Ellis (2010) emphasized two influential categories, learner and social/situational factors suggesting that age, motivation, learning style, gender, task type, etc. could be factors affecting the strategies being used. As highlighted in previous studies, language learning strategy use correlates with learners’ language proficiency supporting the fact that more strategic learners are more successful than less strategic ones (Andersan, 2005; Hong-Nam and Aslan, 2016) and on university level students and academics (Akin, 2016; Akpinar & Çakildere, 2013; Büyükkeles, 2016; Hatipoğlu, 2016; Karataş & Okan, 2019; Külêkçi, 2016; Özmen, 2011; Polat, 2019; Sağlam & Farhady, 2019; Yıldırım, 2010).

Washback may be explained in terms of affective factors in language learning like motivation, self-confidence, learner identity or anxiety (Booth, 2018, p.58). Related literature has revealed the existence of washback of high stakes tests (as student motivation increased) particularly for higher proficiency level students (Buyukkeles, 2016; Pan, 2014; Pan and Newfields, 2013; Reynolds, 2010). Similarly, learner autonomy has been found to be related to the level of self-motivation (Bildi, 2017). As highlighted in Pan’s (2014) study, language learning autonomy was directly affected by the test itself.

Test takers’ preparation strategies for TOEFL iBT were also investigated by some researchers (Liu, 2014; Noviana and Ardi, 2020). While some studies on different high takes tests like GMATE (General Multi-media Assisted Test of English) showed a washback on the students’ preparation practices (Gennaro, 2017), others did not find any washback on skill practices of students who needed to take GEPT (General English Proficiency Test) and TOEIC (Test of English for International Communication) (Pan, 2014) or observed a modest change in their test preparation practices for a proficiency test (Pan and Newfields, 2013). On the other hand, a negative washback was reported by Sağlam and Farhady (2019) who looked into the potential washback effect of the University English Language Proficiency Test prepared locally by a Turkish university. The results showed that the practiced learning strategies focused more on the skills tested in the exam.

In brief, although extensive research has been carried out on washback effect of different exams, comparative studies between local and international high stakes tests are still scarce. To meet these objectives, the following research questions were addressed in this study:

1. Is there any difference in the motivation level of the DDP and UPP students across different proficiency levels?
2. Is there any difference in the learning autonomy level of the DDP and UPP students across different proficiency levels?
3. What are the perceptions of the two groups about the language learning strategies they use across different proficiency levels?

4. Methodology

High-stakes tests are described as the tests ‘whose results are seen – rightly or wrongly – by students, teachers, administrators, parents, or the general public, as being used to make important decisions that immediately and directly affect them’ (Madaus, 1988, p. 7). Having this much of a power, high-stakes tests have serious impacts on test takers and the education program (Alderson and Wall, 1992; Shohamy, 2017). In this study, the global TOEFL iBT and the local English Proficiency Exam were both considered as high-stakes tests as both are used as proofs of language proficiency to allow students to continue their undergraduate studies.

In this study, an exploratory sequential mixed method design was used to explore the washback effect of TOEFL iBT and the English Proficiency Exam on test-takers’ motivation, learner autonomy and test-preparation strategies. In a mixed method research the researcher mixes (or integrates or links) the two forms of data concurrently by combining them (or merging them), sequentially by having one built on the other, or embedding one within the other” (Creswell and Plano Clark, 2011, p. 5). Therefore, in this study, the quantitative data were complemented using qualitative evidence to provide deeper understanding of the findings.

4.1. Setting and participants

This research was approved by Bahcesehir University review board. The study was conducted in an English Language Preparatory School running two different preparatory programs namely, University Preparatory Program (UPP) and Dual Degree Program (DDP). The programs
were run at an English Medium Instruction (EMI) university in Turkey with the minimum requirement of A2 level of proficiency expecting students to engage in routine conversation using basic language skills and strategies. However, to follow their undergraduate courses in various disciplines such as Engineering and Architecture, students need to reach the B2 level of CEFR benchmarks as the courses are in English and the students need to have developed their language skills and strategies as listening to lectures, taking notes or writing academic reports. Apart from the language itself, motivation and autonomy are important predictors for students to pass the exams and start their undergraduate programs.

The two programs offered at the university lead to a different qualification because they prepare students for different tests. DDP works like a foundation program in which students need to pay a tuition. The students who are registered to this program come from economically higher-class families. They take courses both in this university and their partner universities in the USA, and they receive diplomas from both of their universities. Therefore, they need to prove their English proficiency level through a recognized English test like TOEFL iBT or Pearson Test of English (PTE) depending on their program requirements; however, the present program was meant to prepare them for TOEFL iBT.

The students in the UPP program, which is the standard English preparatory program, are supposed to pass the English Proficiency Exam prepared by the university itself. The exam consists of two main sections: 15 items for Reading and 15 for Use of English (section 1), 15 items for Listening, and one essay for Writing (section 2). The students who receive the expected grade from the first section (60 out of 100) have the right to take the second section. The overall passing score is 70 out of 100. Although many of these students start the preparatory school with the aim of learning English, their goal shifts towards passing the proficiency exam since they are not allowed to start their undergraduate studies until they pass the global TOEFL iBT exam. Both UPP and DDP programs are based on an Integrated Skills Based Approach aligned with the Common European Framework of Reference (CEFR) for Languages aiming to promote and support the language development of learners.

In this study, 65 students from DDP program and 87 students from UPP program (152 in total) participated on a voluntary basis. The proficiency levels were defined in accordance with the CEFR (https://www.cambridgeenglish.org/exams-and-tests/cefr/). The number of the participants across proficiency levels in the DDP program was as following: 17 A2 level, 23 B1 level and 25 B2 level. As for the UPP program, there were 29 students involved in the same proficiency levels, A2, B1 and B2. Additionally, 14 students from DDP and 13 from UPP took part in the interviews. All of the participants were of Turkish nationality with an age range between 18 to 19 years.

4.2. Data collection instruments

The present study utilized both quantitative and qualitative data collection instruments. As quantitative data collection instruments, a Motivation Questionnaire and an Autonomous Learning Scale were used whereas the qualitative data were gathered from student interviews.

4.2.1. Motivation Questionnaire

A four-point Likert type motivation questionnaire by Schmidt et al. (1996) was used to measure the motivation level of the participants and examine if any difference existed between the motivation of the two groups of participants across proficiency level. The original questionnaire included three main parts as motivation, preferences for instructional activities, and learning strategies. For the purposes of this study, only the motivation part was implemented due to the fact that the full questionnaire measured the needs, preferences and reasons of learners about learning a foreign language in addition to tapping motivation. The questionnaire comprised 50 items in total. The only change was made in item 9, which mentioned ‘Egypt’ as hometown where the word ‘Turkey’ replaced ‘Egypt’ (See Appendix A). Before administering the questionnaire to the participants, we piloted it with 31 randomly selected students and the Cronbach-alpha reliability score was found to be .88 indicating high reliability of the whole instruments (Cronbach, 1951).

4.2.2. Learning Autonomy Scale

As another data collection tool of this study, Learning Autonomy Scale was adopted from Macaskill and Taylor (2010). The scale was a five-point Likert type scale including 14 items ranging from how much a person enjoyed the learning experiences (subscale 1: independence of learning) to how good they were at planning their times (subscale 2: study habits) (See Appendix B). The scale was piloted with 31 students and the alpha coefficient was found to be .83 and .71 for the subscales and .81 for the whole scale indicating a high reliability (Cronbach, 1951).

4.2.3. Student interviews

Apart from the quantitative tools, semi-structured interviews were carried out with 27 (DDP = 14, UPP = 13) volunteer students. The questions were adapted from the survey prepared by Liu (2014), who explored the relationship between test preparation strategies and the TOEFL iBT scores of the language learners. The scale included general language learning strategies related to each skill such as reading, writing, listening and speaking. Similarly, the interview questions in this study addressed the perceptions of the students enrolled in the two programs in regard with the use of learning strategies related to four language skills. For example, the interviewees were asked how they practiced each skill and why they preferred to use such strategies. The aim of asking such questions was to gain a deeper understanding of the use of language learning strategies of each group.

4.3. Procedures

In this study, the quantitative data were collected from both groups two weeks before the exam as previous studies stated an increase in the exam preparation as the time to take tests approached (Tsagari, 2007). The interviews were conducted during the same week with the questionnaires to gain deeper understanding about the language learning strategies preferred by the two groups of participants.

4.4. Data analysis

This study made use of both qualitative and quantitative data collection tools. To decide on the most suitable analysis techniques, first, the normality of the data was checked. More specifically, Kolmogorov-Smirnov test was employed to test the normality of the whole data obtained from DDP and UPP (see Table 1).

According to the results reported in the table above, motivation and autonomy levels were not normally distributed for DDP, motivation (df (87) = .118, p > .05) and autonomy (df (87) = .119, p > .05) and for UPP, motivation (df (65) = .086, p > .05) and autonomy (df (65) = .100, p > .05).

As the results of the normality tests might be affected by the number of the participants, descriptive statistics were calculated to see whether the data were normally distributed across different proficiency levels of DDP and UPP (See Tables 2 and 3).

| Kolmogorov-Smirnov normality analysis. |
|----------------------------------------|
| Kolmogorov-Smirnov                        | df | Sig. |
| DDP Motivation                           | .086 | 65 | .200 |
| Autonomy                                | .100 | 65 | .175 |
| UPP Motivation                           | .118 | 87 | .168 |
| Autonomy                                | .119 | 87 | .007 |

\*P < .05.
5. Results

Although the gap between mean, median and mode was not meaningfully high (see Table 2), not all of the Skewness and Kurtosis values of the different proficiency levels for motivation and autonomy are between the accepted values of -1 and +1 (see Table 3). To this end, non-parametric tests namely, Kruskal Wallis and Mann-Whitney U tests were implemented (Larson-Hall, 2010). Kruskal-Wallis Test was implemented to understand whether the students’ level of motivation and autonomy change across different proficiency levels whereas Mann-Whitney U test was adopted to compare perceived motivation and perceived autonomy of the DDP and UPP students.

Finally, the qualitative data were gathered from student interviews which were conducted in Turkish to ensure that the participants feel comfortable while expressing their viewpoints. The interview data were analyzed through content analysis (Glaser and Strauss, 1967). Firstly, the data went through open coding before further grouping the emerging categories depending on their similarities. To ensure inter-rater reliability, two experts of English Language Teaching (ELT) analyzed the data and inter-rater reliability was found to be .86 showing an agreement between the raters (Mackey and Gass, 2005).

5. Results

The present study aimed to investigate and compare the washback effect of two high stakes tests, TOEFL iBT and a local English Proficiency Exam, on students’ motivation and their learning autonomy levels and to examine if any difference exists between these two across proficiency levels. Finally, the perceptions of the DPP and UPP students about their use of learning strategies was explored.

First, to examine whether any difference existed between the motivation level of the two groups, the Mann Whitney U test was utilized. According to the obtained results, no significant difference was detected between the motivation levels of DDP students with a mean rank of 81.45 and the UPP students with a mean rank of 72.80 \( (U = 2.505, p = .230) \) (See Table 4).

Furthermore, to find out whether there was any significant difference in the students’ motivation across different proficiency levels in the DDP, the Kruskal Wallis analysis was applied. As reported in the Table 5 below, no significant difference was found across different proficiency level for DDP with a mean rank of 38.16 for A2 students, 43.72 for B1 students and 50.12 for B2 students \( (H (2) = 3.267, p = .195) \).

In addition, there was no significant difference in the motivation level across proficiency level in UPP with a mean rank of 40.29 for A2 students, 29.11 for B1 students, and 31.62 for B2 students \( (H (2) = 3.644, p = .162) \).

Apart from the motivation level of the two groups, this study looked into autonomy level of the DDP and UPP students within and between the two programs across proficiency level. The results obtained from the Mann-Whitney U test showed no significant difference \( (U = 2.478, p = .192) \) between the students studying at Dual Diploma Program (DDP) and University Preparation Program (UPP) programs in terms of their autonomy. The mean rank for DPP students was 81.87 while it was 72.39 for the UPP students (see Table 6).

Finally, for the difference in the two programs related to the autonomy level across the proficiency level, the results of the Kruskal Wallis test indicated no significant difference in the autonomy level of B1 students in UPP and DDP groups \( (U = 382.500, p = .365) \) with mean ranks of 28.19 for UPP and 24.37 for DDP group, respectively. Similar results were revealed about autonomy level of B2 students in UPP and DDP \( (U = 287.500, p = .191) \) with mean ranks of 28.26 and 26.62, respectively, showing no significant difference. However, the autonomy level \( (U = 159.500, p = .047) \) of DDP A2 students was significantly higher than UPP A2 students with mean ranks of 20.50 and 28.6 (See Table 7).

Furthermore, the last research question in this study attempted to find out the perceived learning strategies of the two groups while preparing for the TOEFL iBT and the local English Proficiency Exam. To reveal the strategies of each group, the themes emerging from student interviews were categorized under the common themes as follows: As illustrated in the table above (Table 8), the DDP students, regardless of their proficiency level stated that they prepare for the TOEFL iBT exam by using strategies such as reading books, watching TV programs and talking to foreigners. They agreed that reading books and watching TV programs particularly help them to expand their vocabulary and grammar. Considering these findings, the DDP students said:

### Table 2. Descriptive statistics for proficiency levels.

| N  | Motivation | Autonomy |
|----|------------|----------|
|    | M Med. Mode | M Med. Mode |
| DDP A2 | 17 139 139 134 | 52 55 56 |
| DDP B1 | 23 132 132 116 | 50 50 49 |
| DDP B2 | 25 134 134 129 | 53 53 50 |
| UPP A2 | 29 131 131 131 | 50 50 51 |
| UPP B1 | 29 131 131 131 | 51 52 54 |
| UPP B2 | 29 135 134 129 | 50 50 50 |

*p < .05.

### Table 3. Skewness and Kurtosis normality analysis.

| N  | Motivation | Autonomy |
|----|------------|----------|
|    | Skewness Kurtois | Skewness Kurtois |
| DDP A2 | 17 .020 .283 | 2.246 1.339 |
| DDP B1 | 23 .300 .929 | 1.570 .533 |
| DDP B2 | 25 .316 1.321 | .664 .437 |
| UPP A2 | 29 .225 .497 | -.214 -.517 |
| UPP B1 | 29 -.611 1.298 | .836 -.756 |
| UPP B2 | 29 .026 .627 | 1.30 1.850 |

*p < .05.

### Table 4. The difference between the motivation level of DDP and UPP students.

| N  | Mean Rank | Mann-Whitney U | p  | SE |
|----|-----------|----------------|----|----|
| DDP | 65 81.45   | 2.505          | .230 | 268.304 |
| UPP | 87 72.80   |                |    |    |

*p < .05.

### Table 5. The difference between the motivation level among DDP and UPP students across proficiency level.

| N  | Prof. level | Mean Rank | H  | df | p  |
|----|-------------|-----------|----|----|----|
| DDP | A2          | 38.16     | 3.267 | 2 | .195 |
|    | B1          | 43.72     |      |    |    |
|    | B2          | 50.12     |      |    |    |
| UPP | A2          | 40.29     | 3.644 | 2 | .162 |
|    | B1          | 29.11     |      |    |    |
|    | B2          | 31.62     |      |    |    |

*p < .05.

### Table 6. The Difference between the DDP and UPP Students in terms of Autonomy Level.

| N  | Mean Rank | Mann-Whitney U | p  | SE |
|----|-----------|----------------|----|----|
| DDP | 65 81.87   | 2.478          | .192 | 267.766 |
| UPP | 87 72.49   |                |    |    |

*p < .05.
“One of the strategies I most commonly use to prepare for the TOEFL iBT is reading books and watching TV programs. I learn a lot of vocabulary, and I improve my grammar”. (DPP Student 4, B2 level, Interview data)

“I believe that reading books and watching TV programs helps me learn English. I can practice new words and grammatical rules (DPP Student 13, B1 level, Interview data)

“I prefer to study for the TOEFL iBT reading books as well as watching TV programs. It is very helpful to learn new words and grammar.” (DPP Student 7, B2 level, Interview data)

Additionally, the DPP students agreed that “talking to foreigners” is effective for improving their speaking skills. They highlighted the importance of speaking with foreigners to help them with their accuracy and fluency as illustrated below:

“Practicing speaking is very important before taking the TOEFL iBT. I love talking with foreigners which helps me speak correctly and fluently.” (DPP Student 10, B2 level, Interview data)

“My biggest assistant while preparing to the exam is talking to foreigners who live in my neighborhood. This helps me improve my grammar and feel more comfortable while speaking.” (DPP Student 2, B1 level, Interview data)

“I prefer to talk to the foreigners in my school or neighborhood to practice speaking before I take the TOEFL iBT exam. This is very helpful to speak correctly and fluently.” (DPP Student 9, A2 level, Interview data)

Apart from the common learning strategies expressed by the DPP students, there were some differences of strategy use across proficiency level. First, the B2 level students perceived “note taking” as an effective strategy which helped them with their comprehension as displayed in this excerpt:

“I practice note taking a lot before the exam. I have to understand and be able to take the relevant notes. Thus, I believe that taking notes is an effective strategy to use. (DPP Student 4, B2 level, Interview data)

As for the B1 level DPP students, ‘using online dictionaries’ was considered as an effective strategy to learn and practice vocabulary as shown in this quotation:

“I prefer to use online dictionaries. It is very effective to expand my vocabulary.” (DPP Student 6, B1 level, Interview data)

Finally, the A2 level DPP students emphasized the importance of ‘memorizing words’ to expand their vocabulary as supported in this comment:

“I believe that memorizing words is very important. I use this strategy which helps me improve my vocabulary. (DPP Student 11, A2 level, Interview data)

Similarly, the common themes induced from the interviews of the UPP group were parallel to those of the DPP group except from one theme related to ‘playing computer games’. Specifically, the UPP students regardless of their proficiency levels agreed that watching TV programs, reading books and talking to foreigners are effective strategies. In line with these findings, the UPP students shared the following opinions:

“I prefer to watch programs on TV and read books. I also try to talk to the foreigners to practice speaking. These are effective strategies.” (UPP Student 1, B2 level, Interview data)

“I like reading books and watching programs in English. When I have a chance, I like talking to foreign students in my university as well. (UPP Student 7, B1 level, Interview data)

“I watch TV programs and read some books in English. I also try to talk to my foreign friends. (UPP Student 12, A2 level, Interview data)

Apart from the three common themes shared by the two groups, the UPP group mentioned that ‘playing computer games’ was another strategy they prefer to use particularly to practice grammar and vocabulary. Considering this finding, they shared the following comments:

“I like playing computer games. I think it is very helpful to practice vocabulary as well as grammar.” (UPP Student 12, B2 level, Interview data)

“Playing computer games is very practical to study words and grammar. I enjoy it. (UPP Student 5, B1 level, Interview data)

“I play computer games. It is very effective in terms of studying vocabulary and grammar. (UPP Student 8, A2 level, Interview data)

Moreover, the UPP students showed some differences in their strategy use across proficiency level. For example, the B2 students perceived the ‘use of online’ apps such as Kahoot and Quizlet to be effective applications particularly to practice vocabulary as displayed in the following excerpts:

“I believe that online apps are very useful to practice English. I use Kahoot a lot to revise the words we have learned in the class.” (UPP Student 10, B2 level, Interview data)

“I prefer to practice vocabulary using online apps as Quizlet which helps me practice words”. (UPP Student 4, B2 level, Interview data)

Finally, the B1 level and A2 level UPP students perceived ‘practicing vocabulary’ to be an important learning strategy. They stated that this particular strategy is useful especially for their understanding as expressed in these comments:

Table 7. The difference between the autonomy level of DDP and UPP students across proficiency level.

| Autonomy Level | N   | Mann-Whitney U | p     | SE  | Mean Rank |
|----------------|-----|---------------|-------|-----|-----------|
|                |     |               |       |     |           |
| DDP            |     |               |       |     |           |
| B1             | 52  | 382.500       | .365  | 54.139 | 28.19     | 24.37     |
| B2             | 54  | 287.500       | .191  | 57.390 | 24.91     | 30.50     |
| A2             | 46  | 159.500       | .047* | 43.752 | 20.50     | 28.62     |

*p < .05.

Table 8. The Common Themes emerging from DDP and UPP Students.

| Students | Common themes | DDP | UPP |
|----------|---------------|-----|-----|
|          | reading books, watching TV programs, talking to foreigners | watching TV programs, playing computer games | |
| B2       | watching TV programs, talking to foreigners, reading books, taking notes | watching TV programs, reading books, talking to foreigners, playing computer games |
| B1       | reading books, talking to foreigners, watching TV programs, using online dictionaries | watching TV programs, reading books, playing computer games, talking to foreigners, practicing vocabulary |
| A2       | watching TV programs, reading books, talking to foreigners, memorizing vocabulary | watching TV programs, reading books, playing computer games, talking to foreigners, practicing vocabulary |
“I like practicing vocabulary. I try to learn as many words as possible which helps me to understand better. (UPP Student 8, B1 level, Interview data)

“I prefer to practice vocabulary. If we know different vocabulary, we can understand better (UPP Student 3, A2 level, Interview data)

6. Discussion

The present study aimed to investigate the washback effect of two high stakes tests, a global TOEFL iBT and a local English Proficiency Exam, on students’ motivation and autonomy across different proficiency levels. The study also tried to find out the language learning strategies the students use with regard to their proficiency level and find out the reasons behind their preferences.

According to elicited data, the findings revealed that neither DDP participants, who were supposed to take TOEFL iBT, nor UPP students, who were to take the local Proficiency test, considered themselves as motivated language learners. Similarly, variety across proficiency level did not lead to differences in motivation in any group. These results may imply that the mentioned tests did not have any washback on language learning motivation of students at different proficiency levels. This result contradicts the findings of some researchers (Pan and Newfields, 2013), who investigated the washback of different local University Proficiency exams and found a good amount of positive washback on students’ motivation, especially for the higher proficiency levels (Buğüktekeş, 2016). Similarly, it contradicts the findings of Reynolds (2010) who revealed a negative washback of TOEFL iBT on students with a higher proficiency level.

Some other studies carried out in Turkey have reported a negative washback effect of some nationwide English proficiency exams (Akın, 2016; Akpanar and Çakoldère, 2013; Hatipoğlu, 2016; Kuşçu, 2016; Külekçi, 2010; Özmen, 2011; Polat, 2019; Sayın and Aslan, 2016; Yıldırım, 2010), findings which were not substantiated in our research. However, it should be noted that the exams investigated by previous researchers did not include any productive skills, so they may not be considered quite similar either to the TOEFL iBT or the local English Proficiency exam in question, except for all being high-stakes tests.

Similar to the findings about motivation, the results related to student autonomy showed no washback of the local English Proficiency exam and TOEFL iBT except for DDP A2 students who had higher level of autonomy than their UPP counterparts. This finding may reveal some possible possible washback of the TOEFL iBT for the lower proficiency level students although no possible washback was observed of the local Proficiency test. These results for TOEFL iBT contradict those of Pan (2014), who found an impact on students’ autonomous learning for TOEIC practice as well as findings by Noviana and Ardi (2020), who reported higher learner autonomy in TOEFL test takers.

Finally, the study looked at the perceptions of the students enrolled in the two programs regarding their use of language learning strategies. The findings showed common strategies used by the students of the DDP and UPP programs. The common strategies used by the two groups were related to reading books, watching TV and talking with foreigners which were perceived to be helpful for practicing vocabulary, grammar and use of language. Considering this finding, it might be inferred that reading books and watching TV were perceived as effective strategies which helped the participants of the two programs improve their comprehension of the target language. Besides, talking to the foreigners was also highlighted as an important strategy to practice the use of the language. As there were English-speaking foreign students at the campus studying at various disciplines, the UPP and DPP groups preferred to talk with them to improve their speaking skills.

On the contrary, the only difference between the two groups was that the UPP students preferred to ‘play computer games’ while practicing grammar and vocabulary. The reason behind this finding might be the fact that playing computer games is particularly helpful for practicing grammar and vocabulary which were components of the UPP exam.

When the DPP and UPP students were compared across proficiency level, there were some perceived similarities and differences. Considering the DPP group, the B2 students reported ‘note taking’ to be an important strategy for their comprehension. Additionally, the B1 students expressed ‘using online dictionaries’ whilst the A2 students mentioned ‘memorizing vocabulary’ as effective strategies for practicing and expanding their vocabulary. On the contrary, the B2 students in the UPP highlighted the importance of ‘using online apps’ to revise the target words. These results are in line with the increase in using technological tools to improve the language learning process and to help learners practice the language both in and out of the classroom (Ahmadi, 2018). Finally, the UPP students at B1 and A2 levels perceived ‘practicing vocabulary’ to be an effective strategy for expanding their knowledge on target words. These findings confirm the effectiveness of using learning strategies to achieve learner’s objectives (Woolfolk, 1998) as well as provide support for their use to facilitate the language learning process (Cohen and Weaver, 2006; Ozer, 2002).

7. Pedagogical implications, limitations and suggestions for further research

The present study affords some pedagogical implications for practitioners and test designers. First, although motivation was not affected among the two groups with different proficiency levels, it still needs to be addressed by practitioners and material designers. While tests have the power to affect motivation, the opposite is also true: more motivated test-takers will do better on tests regardless of their posterior effects. Motivating students during classroom learning and the test preparation period is crucial in terms of their overall achievement. Therefore, regardless of the program and proficiency level, students should be provided with materials and engaged in practices that will enhance their motivation while learning a language and preparing for an exam. Similarly, practices on increasing learning autonomy could also be beneficial for language learners. Engaging learners in autonomy fostering activities such as doing research on their topic of interest and preparing a presentation on their favorite topic might help them become more autonomous as well as practice the language skills and strategies more effectively. In addition, students in both programs may be engaged more in practicing learning strategies such as note taking, scanning or making inferences to help them improve their language skills and prepare for the exam as well. Finally, program developers and teacher trainers may offer training programs about effective preparation strategies before taking language exams.

Considering the gathered results, the study has some limitations that could be addressed in further research. First, as the study was conducted at a single university, future research can be expanded by examining and comparing similar language exams at different universities within the same country or across countries. Secondly, a follow up study could be done by implementing a needs analysis, to understand whether there is a gap between the needs of the learners and their current strategy practices they need to improve before taking such exams. One further limitation that could be addressed in future washback studies is to measure motivation and autonomy both at the start of the exam-preparation course and at the end to empirically investigate whether the program has led to any changes in these contracts. Finally, the washback effect of tests on teachers may be further examined to provide deeper understanding about their experiences as well as challenges.

8. Conclusion and recommendations for future research

The results of this study are prominent in terms of addressing student motivation, autonomy and language learning strategies while preparing both for global and local language exams. Although the findings revealed
no differential washback of either test, it seems that teachers play a significant role in determining the washback effect. As teachers are responsible for helping students prepare for and pass the exams, they need to be aware of the needs and preferences of their students and adapt their practices accordingly. Teachers should learn more about effective methodologies to enhance students’ learning, help them become more motivated and autonomous learners as well as promoting positive washback or hindering negative washback in their classrooms. Therefore, training courses should be designed for teachers (particularly on language testing and teacher assessment literacy) as they have tremendous power to motivate students to learn, and to teach them how to prepare for tests.

Declarations

Author contribution statement

Karim Sadeghi, Aylin Balldağ, Enisa Mede: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Funding statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability statement

Data included in article/supplementary material/referenced in article.

Declaration of interests statement

The authors declare no conflict of interest.

Additional information

Supplementary content related to this article has been published online at https://doi.org/10.1016/j.heliyon.2021.e08135.

References

Ahmadi, M.R., 2018. The use of technology in English language learning: a Literature Review. Int. J. Res. Educ. Educ. 3 (2), 115–125.
Akþinar, K.D., Çakildere, B., 2013. Washback effects of high-stakes language tests of Turkey (KPDS and ÜDS) on productive and receptive skills of academic personnel. J. Lang. Linguist. Stud. 9 (2), 81–94.
Akın, G., 2016. Evaluation of national foreign language test in Turkey. Asian J. Educ. Res. 4 (3), 11–21.
Alderson, C., 2004. Foreword. In: Cheng, L., Watanabe, Y., Curtis, A. (Eds.), Washback in Language Testing: Research Contexts and Methods. Lawrence Erlbaum Associates, pp. ix–xii.
Alderson, J.C., Hamp-Lyons, L., 1996. TOEFL preparation courses: a study of washback. Lang. Test. 13, 280–297.
Alderson, J.C., Wall, D., 1992. Does washback exist? [Paper presentation]. In: Symposium on the Educational and Social Impacts of Language Tests: Language Testing Research Colloquium. Lancaster University. Retrieved from: https://eric.ed.gov/?id=EJ945915.
Alderson, J.C., Wall, D., 1993. Does washback exist? Appl. Linguist. 14, 115–129.
Anderson, N.J., 2005. 12 strategy research. In: Hinkel, E. (Ed.), Handbook of Research in Second Language Teaching and Learning. Lawrence Erlbaum Associates, Mahwah, NJ, pp. 782–772.
Bailey, K.M., 1996. Working for washback: a review of the washback concept in language. Lang. Test. 15, 257–279.
Bailey, K.M., 2005. Practical English Language Teaching: Speaking. McGraw-Hill, New York.
Barnes, M., 2010. The Washback of the TOEFL IBT in Vietnam. Unpublished doctoral dissertation. University of Melbourne.
Barnes, M., 2016. The washback of the TOEFL IBT in Vietnam. Austral. J. Teacher Educ. 41 (7), 157–174.
Bildi, S., 2017. Collaborative Learner Autonomy. Springer.
Booth, K.D., 2018. The Sociocultural Activity O High Stakes Standardized Language Testing: TOEIC Washback in a South Korean Context. Springer.
Brown, H.D., 2004. Language Assessment: Principles and Classroom Practices. Pearson Education, White Plains, NY.
Büyükkeles, G., 2016. The Washback Effect of a High-Stakes Exit Test on Students’ Motivation in a Turkish Pre-university EFL Preparatory School. Unpublished masters’ dissertation. University of Reading, England.
Chamot, A.U., 1998. Teaching Learning Strategies to Language Students. Washington. Retrieved from: https://files.eric.ed.gov/fulltext/ED433719.pdf.
Cheng, L., 2005. Changing Language Teaching through Language Testing: A Washback Study. Cambridge University Press, New York.
Cheng, L., Curtis, A., 2004. Washback or backwash. In: Cheng, L., Watanabe, Y., Curtis, A. (Eds.), Washback in Language Testing: Research Contexts and Methods. Erlbaum Associates, pp. 5–18.
Cohen, A.D., 1998. Strategies in Learning and Using a Second Language. Addison Wesley Longman, New York.
Cohen, A.D., Weaver, S.J., 2006. Styles-and-strategies-based Instruction: A Teachers’ Guide. University of Minnesota, Minneapolis, MN.
Crewell, J.W., Plano Clark, V.L., 2011. Designing and Conducting Mixed Methods Research, second ed. Sage Publications, Inc.
Cronbach, L.J., 1951. Coefficient alpha and the internal structure of tests. Psychometrika 16, 297–334.
Deci, E.L., Ryan, R.M., 1985. Intrinsic Motivation and Self-Determination in Human Behavior. University of Rochester, Rochester, New York, Plenum.
Dorney, Z., 2005. The Psychology of the Language Learner: Individual Differences in Second Language Acquisition. Lawrence Erlbaum Associates Publishers.
Dorney, Z., Otto, I., 1998. Motivation in action: a process model of L2 motivation. Work. Paper. Appl. Linguist. 4, 43–69.
Dorney, Z., Ushioda, E., 2013. Teaching and Researching: Motivation. Routledge, New York.
Eccles, J.S., Wigfield, A., 2002. Motivational beliefs, values, and goals. Annu. Rev. Psychol. 53, 109–132.
Ellis, R., 2010. The Study of Second Language Acquisition, second ed. Oxford Publishing, Fox, J., Cheng, L., 2015. Walk a mile in my shoes: stakeholder accounts testing experience with a computer-administered test. TESL Can. J. 32 (9), 65–86.
Gardner, R.C., 1985. Social Psychology and Second Language Learning: the Role of Attitudes and Motivation. Edward Arnold Publishers.
Germann, J.A.D., 2017. The Washback Effects of an English Exit Exam on Teachers and Learners in a Korean university English Program. Unpublished doctoral dissertation. University of Exeter, England.
Glaser, B.G., Strauss, A.L., 1967. The Discovery of Grounded Theory. Aldine.
Graham, P., 2005. Second language learning strategies and Japanese: does orthography make a difference? System 33, 327–339.
Hatipoglu, C., 2016. The impact of the university entrance exam on EFL education in Turkey: pre-service English language teachers’ perspective. Proced. Soc. Behav. Sci. 232, 136–144.
Hong-Nam, K., Leavell, A., 2006. Language learning strategy use of ESL students in an intensive English learning context. System 34, 399–415.
Hughes, A., 1989. Testing for Language Teachers. Cambridge University Press.
Hughes, A., 1993. Backwash and TOEFL 2000. Unpublished manuscript. University of Reading, England.
Karatas, T.O., Okan, Z., 2019. The power of tests in Turkish context: a critical study. J. Lang. Linguist. Stud. 15 (4), 210–230.
Küleþci, E., 2016. A concise analysis of the Foreign Language Examination (YDS) in Turkey and its possible washback effects. Int. Online J. Educ. Teach. 3 (4), 303–315.
Kılıçkaya, F., 2016. Washback of a high-stakes exam on lower secondary school English teacher’s practices in the classroom. Lublin Stud. MODER. LANG. LITETR. 40 (1), 116–134.
Larson-Hall, J., 2010. A Guide to Doing Statistics in Second Language Research Using SPSS. Routledge.
Liu, D.J., 2014. Investigating the relationship between test preparation and TOEFL IBT performance. ETS Res. Rep. Series 2014 (2), 1–13.
Macaskill, A., Taylor, E., 2010. The development of a brief measure of learner autonomy in university students. Stud. High Educ. 33 (5), 351–359.
Mackey, A., Gass, S.M., 2005. Second Language Research: Methodology and Design. Lawrence Erlbaum.
Madaus, G.F., 1988. The influence of testing on the curriculum. In: Tanner, L.N. (Ed.), Critical Issues in Curriculum: Eighty-Seventh Yearbook of the National Society for the Study of Education. University of Chicago Press, pp. 83–121.
McKENzie, R.M., 2008. Social factors and non-native attitudes towards varieties of spoken English: a Japanese case study. Int. J. Appl. Ling. 18 (1), 63–88.
Menken, K., Shohamy, E., Isar, G., 2017. High-stakes tests as de facto language education policies. In: Stephen, M. (Ed.), Language Testing and Assessment, third ed., 10th ed., Springer, pp. 385–398.
Noviana, N., Ardi, P., 2020. Investigating learner autonomy in a TOEFL preparation program. J. English Educ. Stud. (JIES) 3 (1), 10–20.
Oxford, R., Crookall, D., 1988. Research on language learning strategies: methods, findings, and instructional issues. Mod. Lang. J. 73 (4), 404–419.
Ozmen, K.S., 2011. Washback effects of the inter-university foreign language examination on foreign language competencies of candidate academics. Novitas-ROYAL (Res. Educ. Studies) 15 (2), 215–228.
Ozer, B., 2002. Ikiligeirim ve ortaöğretim okullarının eğitim programlarındaki öğrenme stratejileri. Eğitim Bilimleri ve Uygulama 1 (1), 17–32.
Palay铎y, D., 2005. Culture and learner autonomy. In: Palay铎y, D., Smith, R. (Eds.), Autonomy across Cultures: Language Education Perspectives. Palgrave Macmillan, pp. 1–22.
Pan, Y., 2009. A review of washback and its pedagogical implications. VNU J. Sci. Fore. Lang. 25, 257–263.
Pan, Y., 2014. Learner washback variability in standardized exit tests. Electr. J. English Second Lang. 18 (2), 1–30.
Pan, Y., Newfields, T., 2013. Student washback from tertiary standardized English proficiency exit requirements in Taiwan. J. Teach. Learn. 9 (1), 13–29.
Park, G., 2010. Investigation into learning strategies used by effective and less effective EFL learners in Korea. Asian Soc. Sci. 6 (8), 92–104.
Pearson, I., 1998. Test as levers for change. In: Chamberlain, D., Baumgardner, R.J. (Eds.), ESP in the Classroom: Practice and Evaluation. Modern English, London, pp. 98–107.
Polat, M., 2019. Washback effect of Turkish foreign language proficiency exam YDS: a scale development study. J. Lang. Linguist. Stud. 15 (4), 1252–1268.
Purpura, J.E., 1999. Learner Strategy Use and Performance on Language Tests: A Structural Equation Modeling Approach. Cambridge University Press.
Reynolds, 2010. An Exploratory Study of TOEFL Students as Evaluators of ‘washback’ to the Learners. Unpublished master’s dissertation. The University of Queensland, Australia.
Sayın, B.A., Aslan, M.M., 2016. The negative effects of undergraduate placement examination of English (LYS-5) on ELT students in Turkey. Part. Educ. Res. 3 (1), 30–39.
Sagtam, A.L.G., Farhady, H., 2019. Can exams change how and what learners learn? Investigating the washback effect of a university English language proficiency test in the Turkish context. Adv. Lang. Lit. Stud. 10 (1), 177–185.
Schmidt, R., Borale, D., Kansabgy, O., 1996. Foreign language motivation: internal structure and external connections. In: Oxford, R. (Ed.), Language Learning Motivation: Pathways to the new century. University of Hawaii Press, pp. 9–70.
Shohamy, E., 2017. Critical language testing. In: Shohamy, E., Or, I.G., May, S. (Eds.), Language Testing and Assessment, third ed., 10. Springer, pp. 441–454.
Spolsky, B., 2017. History of language testing. In: Shohamy, E., Or, I.G., May, S. (Eds.), Language Testing and Assessment, third ed., 10. Springer, pp. 375–384.
Stoneman, B.W.H., 2006. The impact of an Exit English Test on Hong Kong Undergraduates: A Study Investigating the Effects of Test Status on Students’ Test Preparation Behaviors. Unpublished doctoral dissertation. Hong Kong Polytechnic University, China.
Taylor, L., 2005. Washback and impact. ELT J. 59 (2), 54–155.
Tsagari, D., 2007. Review of Washback in Language Testing: what Has Been Done? what More Needs Doing? Unpublished Manuscript. Lancaster University, England.
Tsagari, D., 2011. Washback of a high-stakes English exam on teachers’ perceptions and practices. In: Selected Papers from the 19th ISTAL. Retrieved from. http://ejournals.lib.auth.gr/thal/article/view/5521/5426.
Watanabe, Y., 2004. Methodology in washback studies. In: Cheng, L., Watanabe, Y., Curtis, A. (Eds.), Washback in Language Testing: Research Contexts and Methods. Lawrence Erlbaum Associates, pp. 19–27.
Woolfolk, A.E., 1998. Educational Psychology. Allyn & Bacon, USA.
Wu, J., Lee, M.C., 2017. The relationships between test performance and students’ perceptions of learning motivation, test value, and test anxiety in the context of the English benchmark requirement for graduation in Taiwan’s universities. Lang. Test. Asia 7 (9), 1–21.
Yıldırım, O., 2010. Washback effects of a high-stakes university entrance exam: effects of the English section of the university entrance exam on future English language teachers in Turkey. Asian EFL J. Quart. 12 (2), 92–116.