| **RESEARCH ARTICLE**

Differentiating Washback Effects Across Education Settings: The Case of a Standardized Test of English Proficiency in China

Liu Weiwei

School of Foreign Studies, South China Normal University, Guangzhou, Guangdong, People's Republic of China
Corresponding Author: Liu Weiwei, **E-mail**: Dr_liuweiwei@126.com

| **ABSTRACT**

The potential washback effects of large-scale public examinations have long been recognized in language testing. Despite the considerable number of studies that have attempted to explore washback effects from various perspectives, very few have attempted to differentiate washback effects of the same test across specific education settings. The present study aimed to fill the research gap by comparing the washback effects of TEM (Test for English Majors), a national standardized English proficiency test in China, in two distinct university settings. 237 students participated in the study and the washback effects of TEM on their learning practice and test perception were examined and subsequently compared. The study identified differential washback effects and concluded that washback should be viewed as a type of contextualized effects. Findings of the study have important implications for the design of language assessment tools to be used in a variety of educational settings.

1. Introduction

Washback is the influence of language testing on teaching and learning (Alderson & Wall, 1993; Hughes, 1989). The increasing recognition of the importance of language assessment in language teaching and learning has resulted in a corresponding increase in studies that explore the washback of language tests in various educational settings. In particular, the washback effects of large-scale standardized tests have received extensive empirical attention. Studies on the washback of internationally accepted proficiency tests of English such as International English Language Testing System (IELTS) and Test of English as a Foreign Language (TOEFL) are abundant in language assessment literature (e.g., Erfani, 2012; Estaji & Ghiasvand, 2019; Sadeghi et al., 2021). Washback effects of nationwide standardized English proficiency tests in China, such as Test for English Majors (TEM), College English Test (TEM), Nationwide Unified Examination for Admissions to General Universities and Colleges (NMET), have also been examined from a considerable variety of perspectives (e.g., Dong et al, 2023; Xu & Liu, 2018; Zhang & Bournot-Trites, 2021). These studies tended to focus on the washback effects of language proficiency tests in single educational settings or on single stake-holder groups. Very few empirical studies have been conducted to explore the potential differences in washback effects of large-scale standardized language proficiency tests across various education settings.

To fill the research gap, the present study compared the washback effects of TEM in two distinct educational settings: a comprehensive university and a specialist university of foreign languages. The following two research questions were addressed:

1. What are the washback effects of TEM on students' test perceptions and language learning practice?
2. Does the washback of TEM vary across educational settings?

2. Literature Review

2.1 Theoretical construct of washback

Hughes (1989) defined washback as the impact of testing on teachers' teaching and learners' learning processes. In language teaching and learning literature, numerous definitions of washback can be found. For example, Shohamy et al. (1996) conceptualized washback as "the connections between testing and learning" (p. 298). Gates (1995) defined washback as "the influence of testing on teaching and learning" (p.101). Shohamy (1992) defined washback as "the utilization of external language tests to affect and drive foreign language learning" and emphasized that "this phenomenon is the result of the strong authority of external testing and the major impact it has on the lives of test-takers" (p. 513). Messick (1996) defined washback as "the extent to which the introduction and use of a test influences language teachers and learners to do things they would not otherwise do that promote or inhibit language learning" (p.241).

Alderson and Wall (1993) proposed 15 hypotheses about washback and specified how tests would exert washback effects on different aspects of teaching and learning. Six hypotheses of Alderson and Wall (1993) dealt with the washback on learning, including learning content, strategy, and depth of learning and learners' attitudes. Another six focused on the impacts on teaching, including teaching content, method and teachers' attitudes. Several hypotheses in this model of washback also addressed the possibility that the washback of testing might vary with the importance of tests and the characteristics of individuals. Based on the 15 hypotheses, Alderson and Wall (1993) subsequently explored the impact of washback of O-level exam in Sri Lanka, an English exam for the intermediate level learners. They classified washback into positive and negative effects. Later on, in a study into the washback of TOEFL, Alderson and Hamp-Lyons (1996) found that both the teaching content and teaching methods of the TOEFL preparatory courses were influenced. They thus revised the 15 hypotheses in the washback model of Alderson and Wall (1993) and proposed a 16th hypotheses, highlighting that both the amount and type of washback can vary across teachers and learners. As a dichotomy model, Alderson and Wall's (1993) washback hypotheses covered the impacts of testing on both teaching and learning, and on both teachers and learners. Yet its focus was limited to the micro aspects of teaching and learning that could be influenced by tests.

Hughes' (1993) proposed a trichotomy model of washback by distinguishing the effects of washback among participants, processes and products. Participants in this model included all those whose perceptions and attitudes towards their work may be affected by the test, such as students, teachers, curricular designers, test-affected material writers, and researchers. Process referred to all the actions by the participants which may contribute to the process of learning such as teaching material development, curriculum design, various teaching methods and content, and the use of test-taking methods. The product referred to what is learned and the quality of the learning such as skills and truth. Furthermore, Hughes (1993) highlighted that the nature of a test might initially affect the perceptions and attitudes of the participants and these perceptions and attitudes would subsequently affect their learning practice, which in the end would affect the products. The washback model of Hughes (1993) extended the view of washback to all stake-holders of language tests and to the whole process, and therefore was often referred to as 3Ps Model of washback.

2.2 Empirical studies on washback of language tests

Studies on washback of language tests have gone through several distinct phases. Early studies on washback trace back to the 1950s when scholars began to recognize the relationship between testing and teaching. Washback research in the early stage focused on verifying the existence of washback and constructing theoretical models. In the wake of the 15 hypotheses proposed by Alderson and Wall (1993), a large body of empirical studies on washback were carried out in a remarkable variety of test settings around the world. Washback research conducted in this phase tended to focus on teaching and the relationship between washback and teaching characteristics such as teaching content, methods and materials. Cheng and Curtis (2012) advocated for empirical studies on the relationship between test performances and learner characteristics such as gender, grade, motivation and anxiety and ushered in a new phase of washback research in which research focus was gradually shifted from the teachers to the learners. In terms of research approach, most studies have attempted to examine washback through questionnaire and observation-based case studies of participants and processes (e.g., Alderson & Wall, 1993; Burrows, 1998; Cheng, 2005a; Ferman, 2004; Ghorbani et al, 2008; Qi, 2004; Shohamy et al. 1996; Wall, 2005; Watanabe, 2004). Furthermore, washback of large-scale language tests, particularly those high-stake tests that are used to screen applicants for educational programs and regulate access to educational opportunities, received extensive research attention in all phases of washback research.

In the early phase of washback research, Watanabe (1996) investigated the washback of the university entrance examination in Japan and examined the effect of the university entrance examination on the prevalent use of the grammar-translation method in language classrooms in Japan. Two teachers from a preparatory school participated in the study and the washback on them was examined through interviews and classroom observations. The research revealed that washback would vary with teacher factors such as educational background, personal beliefs, teaching experience and academic qualifications. Shohamy et al. (1996) examined the washback of two national language tests of Arabic as a second language (ASL) and English as a foreign language (EFL) on three group of stakeholders, teachers, students and language inspectors. Through questionnaires, interviews and

document analysis, the study revealed different washback patterns for the two tests. The EFL test was found to affect teaching activities, time devoted to test preparation and production of new teaching materials. Furthermore, the study demonstrated that washback would vary over time, due to factors such as the status of the language and the uses of the test.

In countries with centralized examination systems where national tests are often used as the primary device through which educational reform and innovation are engineered and competent candidates for education programs are selected, a considerable number of washback studies have been conducted to explore the potential impacts of these high-stake exams. For example, Luxia (2007) explored the washback of the writing test in the National Matriculation English Test (NMET) in China and revealed how the urge to raise scores in the real test situation shaped language teaching and learning practice. Participants in the study consisted of an extensive sample of the NMET stake-holders, comprising test constructors, teachers, and students. Data collection instruments included interviews, classroom observations, and questionnaires. The results showed discrepancy between the actual and the anticipated effects of writing test of NMET on the teaching of EFL writing in secondary schools in China. Although communicative features were highlighted in the NMET writing task, such features were not observed in the school practice that prepared students for this task. Both teachers and learners neglected the communicative context of writing while emphasizing the testing situation and the assumed preferences of the markers. Zhan and Andrews (2014) examined the washback of College English Test Band 4 (TEM-4) on Chinese non-English-major undergraduates' out-of-class learning by following three cases in one university from the onset of their college English study to the examination day. A total of 106 diary entries and 30 post-diary interview recordings were collected in the study. Findings of the study suggested that the test exerted more impacts on what students learned than on how they learned and the washback of the test was closely related to students' perception of the test. Nahdia and Trisanti (2019) examined washback of the national English exam in Indonesia. Two ninth-grade teachers and 16 ninth-grade students of a junior high school participated in the study. Both questionnaire and interview were employed in the study which revealed that the national examination exerted both positive and negative washback on the participants. The test enhanced students' language learning motivation. Yet it promoted students to focus their attention merely on contents that were included in the exam.

The washback of large-scale proficiency tests that are administered internationally has attracted much empirical attention. For example, Allen (2016) investigated the washback of the International English Language Testing System (IELTS) Academic exam on learners' test preparation strategies and score gain, and the mediating factors influencing washback when learners in an EFL context are not enrolled in test preparation courses. Two IELTS Tests were administered to 190 undergraduates at a Japanese university over a 1-year period. A survey instrument was employed to collect data about test preparation strategies for both tests. Test scores were compared to assess score gain. Interviews were conducted with 19 participants to investigate the factors mediating washback. Findings of the study suggested that IELTS exerted positive washback on learners' language ability and test preparation strategies, specifically regarding productive skills, which learners in the study context had neglected in their previous language study. Sadeghi et al. (2021) examined the washback effect of TOEFL on students' motivation and learner autonomy. The study also examined whether students' proficiency level moderated the potential washback effect. The study context was two English language preparatory programs offered at a state university in Turkey. The participants of the study were 152 students whose proficiency levels ranged from A2 to B2 on the CEFR Framework. Data collection instruments employed in the study included motivation questionnaire, autonomous learning scale and student interviews. The results revealed that TOEFL exerted impacts on students' language learning strategies though it did not change students' motivation and autonomy.

3. Methodology

3.1 Context and participants

237 undergraduate students from two Chinese universities participated in the present study. 129 participants (54%) were from a comprehensive university that offers undergraduate and postgraduate programs in a wide range of disciplines. 108 participants (46%) were from a specialist university of foreign languages where degree programs centered around foreign languages and international studies. In terms of linguistic background, all participants spoke English as a foreign language. All participants had sat for TEM within the past two years.

Test of English Majors (TEM) is a national English proficiency test tailored for undergraduate students majoring in English in Chinese universities. TEM is administered nationwide by the National Advisory Committee for Foreign Language Teaching (NACFLT) of China. The aim of the test is to identify students' proficiency levels and subsequently examine whether they reached the required levels of English language proficiency specified in the National College English Teaching Syllabus for English Majors (NACFLT, 2000a). The TEM test battery consists of TEM4 and TEM4 Oral, which are administered at the end of the students 2nd academic year, and TEM8 and TEM8 Oral, which are administered at the end of students' undergraduate program. TEM reports scores in four levels, namely, in descending order, "excellent" (composite score 80 or above), "good" (composite score between 70 and 79), "pass" (composite score between 60 and 69) and "failure" (composite score below 60) (NACFLT, 2000b). TEM does not report composite scores or section scores.

All participants had received their TEM results when they participated in the present study. Their TEM results covered the full range of TEM score levels, indicating that all the four score levels of TEM were represented in the present study. 187 participants (79%) obtained "good" in their latest TEM, 23 (9.5%) obtained pass and 19 (8%) obtained "excellent". Eight participants (3%) had obtained "failure". It should be noticed that TEM would issue a certificate to test-takers scoring "pass" or above and the certificate specified test-takers' levels of performance. Those who failed to obtain the certificate can assume that they scored "failure" in TEM.

3.2. Instruments
The present study adopted both quantitative and qualitative research instruments. A questionnaire was administered to all participants. The questionnaire consisted of three parts. The first part collected students' demographic information such as gender, age, educational background and linguistic background. Students were also requested to provide their latest TEM score level in the first part of the questionnaire.

The second part of the questionnaire investigated students' perception of TEM and language tests in general. This part of the questionnaires consisted of 11 items ranging from whether TEM affected students' personal relationships and whether TEM score levels accurately reflected test-takers' proficiency levels to how the TEM has changed their emotional status such as anxiety and sense of achievement and how TEM would influence their employment prospects. The third part of the questionnaire was composed of 10 items and aimed to examine the impacts of TEM on students' learning practice. This part of the questionnaire was adapted from the questionnaire Hung and Huang's (2019) designed for students to report the impacts of a proficiency test on various aspects of language learning and the questionnaire Cheng's (2005b) designed to investigate how Hong Kong College Entrance Exam affected students' attitude towards language learning and changes in their learning practice. All items in the second and third part of the questionnaire were placed on a 5-point Likert scale ranging from "strongly disagree" (1) to "strongly agree" (5).

In addition to the questionnaire, semi-structured interviews were carried out with 27 (11%) of the participants to collect qualitative data that might supplement the quantitative analysis in the present study. The interviews were all conducted in Chinese and participants answered questions concerning their personal experience of preparing for TEM and their perceived changes in language learning both before and after TEM. To ensure that all four score levels of TEM would be represented in the interviews, students from each of the four score levels were invited.

3.3 Procedures
The online questionnaire was distributed through "Wen Juan Xing", an online data collection platform, to English majors of two universities in China. 242 responses were received and 5 were excluded from the analysis due to incomplete or invalid answers. Independent sample t-test was performed to compare the perception of TEM and learning practice of students from the two universities, as reported in the second and third part of the questionnaire respectively. After the quantitative analysis of the data collected from the questionnaire, some participants were invited to participate in face-to-face or online interviews. In the selection of participants for the interviews, the present study ensured that there would be students representing each of the two universities and each of the four score levels.

4. Results and Discussion
4.1 Washback on tests perception
The test adopted a mixed research design to investigate the washback effect of TEM on students' test perception and learning practice. The second part of the questionnaire, which consisted of 11 items, was designed to gain insights into whether and how preparing for and sitting for TEM affected students' perceptions of and attitude towards the test and language tests in general. The 11 items placed on a five-point Likert scale ranged from how much importance they attached to TEM (subscale 1: importance), how TEM and other language tests affect their relationship with teachers and classmates (subscale 2: personal relationship), whether TEM and other language tests had caused anxiety or depression (subscale 3: affection), relationship between TEM and language learning motivation (subscale 4: motivation), to changes in language proficiency level caused by TEM or other language tests (subscale 5: learning outcomes). The questionnaire was piloted with 29 students from the comprehensive university in the present study and the alpha coefficient was found to be .71, indicating a high internal reliability (Cronbach, 1951).

Table 1. Descriptive statistic result on students' perception of TEM and language tests

Item	Mean	S.D.
1	3.82	1.07
2	3.32	1.32
3	3.09	1.28

4	3.6	1.25
5	4.01	1.01
6	3.08	1.35
7	3.52	1.31
8	3.77	1.05
9	3.71	1.01
10	3.73	1.09
11	3.34	1.33

As can be seen in Table 1, participants responded positively to all the 11 items in the second part of the questionnaire. This indicated that from the perspective of the participants, TEM did exert positive impacts on their perception of language testing in general. The highest mean was reported for Item 5: English tests such as TEM are important for my future employment prospects. This suggested that EFL students from both of the two universities attached great importance to their performances on language proficiency tests as they believed that the results were very likely to influence their chance of landing a job shortly after graduating from the university. The second highest mean was reported for Item 1: TEM and other large-scale language tests would affect my personal image both at college and in the future workplace. This again suggested that participants tended to consider language tests as important activities that would exert impacts on their life in the future. In terms of personal relationship (items 8 and 9), participants generally agreed that their relationship with students and teachers would be affected by language tests such as TEM, though they tended to think that the impacts were moderate. Regarding the impacts of TEM and other language tests on participants' motivation of learning English (items 2, 4 and 7), participants from the two universities indicated that language tests would always motivate them to make sustained efforts to improve their language proficiency levels. An interesting finding is that in terms of learning outcome, participants moderately agreed that scores of TEM reflected their genuine language proficiency level (item 11), though they admitted that preparing for TEM or other language tests actually enhanced their language proficiency level.

In the follow-up interview, participants expressed their concerns about the growing importance the Chinese society, particularly the business sector, placed on scores of language proficiency tests. Several participants referred to TEM as "the single most important test" during their four-year university study. Participants generally thought that the preparation for the exam had improved their English language proficiency levels. One of the participants mentioned that "*the test has magically improved my learning efficiency because realizing that an important test is approaching can help me concentrate more on what I am doing and overcome my procrastination, and most importantly force me to read and listen to English more extensively*". Most of the participants listed enriched lexical resources or expanded vocabulary as one of the most prominent benefits they derived from the preparation for TEM because they "*kept an organized vocabulary notebook*" during the preparation for TEM. Since preparation for TEM was generally done on one's own, without much guidance from the teachers or collaboration with fellow students, most participants did not agree that TEM exerted very remarkable impacts on their personal relationships with teachers or classmates. Some explained that the fact TEM reported test-takers' score levels individually and only disclosed score levels that could be associated with specific individuals to university administers meant that students in the same undergraduate program usually did not know each other's score, even after the disclosure of TEM scores. Several participants pointed out that the scores did not reflect test-takers' genuine language ability because almost all the questions on TEM were multiple-choice questions. Guessing could be a prominent factor that potentially affected students' final scores, particularly when they found it difficult to finish all the questions within the stipulated period of time. Multiple-choice questions were considered by some participants as "an acceptable yet annoying" question type that restricted their thinking. As one participant admitted, "*I fully understand that for a large-scale language test with such a big test population, multiple-choice questions would be the most convenient means at the disposal of the test designers. But when a reading question was presented with a defined set of options in TEM, I felt that I was forced to answer in a way that I otherwise would never have*."

4.2 Washback on students' learning practice
In the present study, the washback of TEM on test-takers' learning practice was investigated by the third part of the questionnaire. The 10 items in this part of the questionnaire explored students' learning practice in three aspects: language skills and areas emphasized during the test preparation, learning strategies during TEM preparation and changes in learning practice after TEM. Table 2 presents the descriptive statistic results on students' learning practice.

Table 2. Descriptive statistic results on students' learning practice

Item	Mean	S.D.
1	3.92	1.01
2	1.31	1.13
3	2.43	1.28
4	3.7	1.15
5	3.82	1.07
6	3.28	1.25
7	1.72	1.31
8	3.47	1.15
9	3.51	1.21
10	1.34	1.25

Participants reported the highest mean for Item 1 (mock exam). Almost all participants agreed that they completed a lot of mock exams before they sat for TEM. In terms of the language skills that were emphasized during the test preparation process, very few students spent much time on sharpening reading skills (item 2) and participants devoted the largest proportion of their test preparation time to listening (item 5). Writing (item 3) was the area of learning that most participants spent little time on. Extensive reading (item 7) and watching English movies (item 10) were not considered by many participants as an effective way to enhance their English proficiency level. Nevertheless, participants indicated that they regularly listened to English news (item 8) before TEM to hone their listening skills and continued to do even after TEM (item 9).

Interviews with the participants revealed similar results. Some participants admitted that they were reluctant to read extensively in reading because they believed that test-takers' scores in the reading section of TEM was largely determined by reading strategies or test strategies rather than reading experience. As one participant revealed, "*TEM requires fast reading and even skimming. You have but one goal, that is, to answer all the questions correctly and as fast as possible*". Several participants regarded the writing section of TEM as the easiest section in the TEM battery because they could get at least a moderate score by merely memorizing a well-written template several days before the test date. Most students considered watching English movies as a relaxing hobby rather than an effective way to learn English. Participants generally agreed that they listened to English news regularly while preparing for TEM. Reasons they provided for the lopsided emphasis on listening during TEM test preparation included: 1)listening is the most difficult part in every year's TEM; 2) the news listening task of TEM posed great challenges for test-takers and 3) they had to familiarize themselves with the fast rate of speech and various accents of the news anchors. Several participants explained in the interview that they continued listening to English news regularly even after TEM because it turned out be an integrated language task rather than training in one discrete language skill. As one participant said, "*listening to English news is a very effective approach to improving English language proficiency because you have to draw on your lexical, syntactic and phonetic knowledge when you listen and you can learn new vocabularies, improve your English pronunciation and hone your listening and writing skills all at once*".

4.3 Differential washback effect of TEM
The present study attempted to differentiate washback effects of TEM on test-takers from two educational contexts: a comprehensive university and a specialist institution of foreign languages in China. Table 3 and Table 4 present the t-test statistic results on perceptions and learning practice of participants respectively.

Results of the independent sample t-test suggested that in terms of perception of TEM and language tests in general, participants from the specialist university reported stronger impacts on their motivations (items 2, 4 and 7) and they tended to believe that TEM had motivated them to make more efforts to improve their English proficiency level. While the washback effects of language proficiency tests on EFL learners' language learning motivation have been extensively explored (e.g., Sadeghi et al., 2021; Sardi et al., 2022), little attention has been given to potentially differential washback effects on motivation across various educational settings. Through the comparison of participants' responses to items 2, 4 and 7 in the second part of the questionnaire, the present study identified significant differences in the washback effect of TEM on students from two different universities.

Table 3. T-test statistic result on students' perception of TEM and language tests

Item	T-value	Sig.
1	1.183	0.239
2	1.085	**0.031**
3	1.588	0.115
4	1.804	**0.003**
5	1.834	0.137
6	1.088	0.073
7	1.224	**0.032**
8	1.183	0.239
9	-0.387	0.700
10	1.588	0.115
11	2.837	0.085

Regarding students' language learning practice before and after TEM, the results revealed that students from the specialist university engaged in speaking (item 4) significantly more than students from the comprehensive university did during the preparation for TEM. Yet they spent significantly less time on listening (item 5) than students from the comprehensive university, indicating that the washback effects of language tests on the test preparation activities of students could actually vary across educational settings. It should also be noticed that students from the two universities have different attitudes towards extensive reading (item 7). Students from the specialist university tended to adopt a significantly more favorable attitude towards the role extensive reading can play during the preparation for TEM.

Table 4. T-test statistic results on students' learning practice

Item	T-value	Sig.
1	1.138	0.293
2	-0.378	0.700
3	1.588	0.125
4	2.873	**0.005**
5	-2.802	**0.000**
6	3.081	0.053
7	-3.214	**0.002**
8	1.183	0.239
9	-0.387	0.700
10	1.588	0.115

Qualitative data collected from the interviews confirmed the differential washback effects between the comprehensive university and specialist university that were revealed by the quantitative analysis of the data and provided insights into possible reasons for

the differences in TEM washback across university settings. Although students from both of the two universities attached considerable importance to TEM, several participants from the specialist university of foreign languages mentioned that they felt obliged to obtain a high score in TEM because students from undergraduate programs other than English in their university were allowed to sit for TEM and the majority of these non-English majors were able to pass TEM. Peer pressure seemed to be a contributory factor that led to the sustained efforts students from the specialist university made during their preparation for TEM. In the comprehensive university, on the other hand, TEM were made available exclusively to English majors. The variance of administrative policies of TEM between comprehensive universities and specialist universities in China can partially account for the observed differences in washback effect on motivation. The differences in education tradition and philosophy between the two universities also emerged as salient factors that might explain the other observed differences in washback of TEM. For example, several participants from the specialist university mentioned the mandatory English reading list of the university and recalled how the importance of extensive reading was reiterated throughout their undergraduate education. As regard to speaking skills, students from the specialist university admitted that TEM Speaking was increasingly recognized as an important benchmark of English proficiency in the Chinese society and employers often had higher expectations for graduates of specialist universities of foreign languages.

5. Conclusion

The overarching aim of the present study is to examine differential washback effects of language tests across educational settings. The washback of TEM on the test perception and learning practice of students from two universities was examined with both qualitative and quantitative instruments. Results of the study indicated that TEM exerted washback effects on both students' test perceptions and learning practices. Yet the washback effects were significantly different between the two universities. The present study concluded that washback of a language test should be viewed as a type of contextualized effects that could vary with the specific educational context within which the test is administered or the test-takers prepare for the test.

Based on the findings of the present study, several implications and suggestions can be formulated for EFL teachers, learners and test designers. First, since the study has empirically identified university type as a mediator of washback of language tests, it highlights the need for educational settings to be incorporated in theoretical constructs of washback. Furthermore, test designers and EFL teachers should take the characteristics of the specific educational contexts within which the test will be administered into consideration in the validation of language tests. EFL learners, on the other hand, should be fully aware that the characteristics of the learning context would shape their learning practice both before and after language tests. Finally, large-scale language proficiency tests such as TEM can be an effective tool to improve students' language learning motivation, though the exact washback effects on motivation can be context-specific.

The study is not exempt from limitations. Firstly, all the participants in the present study were English majors at Chinese universities. Despite their differences in educational context, they actually constituted a fairly homogeneous group. Caution should taken when generalizing the findings of the present study to other cultural or linguistic contexts. Furthermore, while the TEM test battery consists of four tests, TEM 4, TEM 4 Oral, TEM 8 and TEM 8 oral, the present study did not distinguish these tests. More empirical studies should be carried out to investigate the differential washback effects of these four tests respectively.

Conflicts of Interest: The authors declare no conflict of interest.
Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References

Alderson, J. C. & Wall, D. (1993). Does Washback Exist?. Applied Linguistics, 14(2): 115-129.
Alderson, J. C., & Hamp-Lyons, L. (1996). TOEFL preparation courses: A study of washback. Language testing, 13(3), 280-297.
Allen, D. (2016). Investigating Washback to the Learner from the IELTS Test in the Japanese Tertiary Context. Language Testing in Asia, 6(1): 7-26.
Burrows, C. J. (1998). Searching for washback: an investigation of the impact on teachers of the implementation into the Adult Migrant English Program of the assessment of the Certificate in Spoken and Written English (Doctoral dissertation, Macquarie University).
Cheng, L. & Curtis, A. 2012. Test impact and washback: Implications for teaching and learning [A].
Cheng, L. (2005a). Changing language teaching through language testing: A washback study. In M. Milanovic & C. Weir (Eds), Studies in language testing. Cambridge, UK: Cambridge University Press.
Cheng, L. (2005b). Changing language teaching through language testing: A washback study (Vol. 21). Cambridge university press.
Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. psychometrika, 16(3), 297-334.
Dong, M., Fan, J., & Xu, J. (2023). Differential washback effects of a high-stakes test on students' English learning process: evidence from a large-scale stratified survey in China. Asia Pacific Journal of Education, 43(1), 252-269.
Dordrecht: Kluwer Academic Publishers.
Erfani, S. S. (2012). A Comparative Washback Study of IELTS and TOEFL iBT on Teaching and Learning Activities in Preparation Courses in the Iranian

Context. English Language Teaching, 5(8), 185-195.

Estaji, M., & Ghiasvand, F. (2019). The washback effect of IELTS examination on EFL teachers' perceived sense of professional identity: Does IELTS related experience make a difference?. Journal of Modern Research in English Language Studies, 6(3), 83-103.

Ferman, I. (2004). The washback of an EFL national oral matriculation test to teaching and learning. Washback in language testing: Research contexts and methods, 191-210.

Gates, S. (1995). Exploiting washback from standardized tests. Language testing in Japan, 101-106.

Ghorbani, M. R., Samad, A. A., & Gani, M. S. (2008). The washback impact of the Iranian university entrance examination on pre-university English teachers. Journal of Asia TEFL, 5(3).

Hughes, A. 1989. Testing for Language Teachers [M]. Cambridge: Cambridge University Press, 44-46.

Hughes, A. 1993. Backwash and TOEFL 2000 [Z].Unpublished manuscript, University of Reading, Reading, UK.

Hung, S. & Huang, D. 2019. Standardized proficiency tests in a campus-wide English curriculum: a washback study [J]. Language Testing in Asia. 9. 21.

Jin, Y., & Fan, J. (2011). Test for English majors (TEM) in China. Language Testing, 28(4), 589-596.

Luxia, Q. (2007). Is testing an efficient agent for pedagogical change? Examining the intended washback of the writing task in a high-stakes English test in China. Assessment in Education, 14(1), 51-74.

Messick, S. 1996. Validity and Washback in Language Testing [J]. Language Testing, 13(3): 241-256.

NACFLT (2004a). Syllabus for TEM4. Shanghai: Shanghai Foreign Language Education Press.

NACFLT (2004b). Syllabus for TEM8. Shanghai: Shanghai Foreign Language Education Press.

Nahdia, K., & Trisanti, N. (2019). Examining washback of national examination: EFL teachers and students' perceptions. In UNNES International Conference on English Language Teaching, Literature, and Translation (ELTLT 2018) (pp. 273-277). Atlantis Press.

Qi, L. (2004). Has a high-stakes test produced the intended changes. Washback in language testing: Research contexts and methods, 171-190.

Sadeghi, K., Ballıdağ, A., & Mede, E. (2021). The washback effect of TOEFL iBT and a local English Proficiency Exam on students' motivation, autonomy and language learning strategies. Heliyon, 7(10).

Sardi, A., Surahmat, Z., & Nur, S. (2022). The Washback of Intensive TOEFL Training Program (ITTP) on Student's Learning Motivation. ELS Journal on Interdisciplinary Studies in Humanities, 5(4).

Shohamy, E. 1992. Beyond Proficiency Testing: A Diagnostic Feedback Testing Model for Assessing Foreign Language Learning [J]. The Modern Language Journal, 76:513-531.

Shohamy, E., Donitsa-Schmidt, S., & Ferman, I. (1996). Test impact revisited: Washback effect over time. Language testing, 13(3), 298-317.

Wall, D. (2005). The impact of high-stakes examinations on classroom teaching: A case study using insights from testing and innovation theory (Vol. 22). Cambridge University Press.

Watanabe, Y. (2004). Methodology in washback studies. In Washback in language testing (pp. 41-58). Routledge.

Watanabe, Y. (1996). Does Grammar-Translation Come from the Entrance Examination? Preliminary Findings from Classroom-Based Research [J]. Language Testing, 13(3):318-333.

Xu, Q., & Liu, J. (2018). A study on the washback effects of the Test for English Majors (TEM). New Springer.

Zhan, Y., & Andrews, S. (2014). Washback effects from a high-stakes examination on out-of-class English learning: Insights from possible self theories. Assessment in Education: Principles, policy & practice, 21(1), 71-89.

Zhang, H., & Bournot-Trites, M. (2021). The long-term washback effects of the National Matriculation English Test on college English learning in China: Tertiary student perspectives. Studies in Educational Evaluation, 68, 100977.