



An Examination of the Washback Effect on Iranian EFL Learners' Reading Comprehension: Any Implications for Text Difficulty in the Classroom?

Shabnam Kurosh Khanshan*

Islamic Azad University, Bonab, Iran

shabnamkorosh@gmail.com

*Corresponding author

Abstract

In order to better understand the mechanisms of washback, the present study aimed to explore the washback effect of the entrance examination in Iran on language learners' achievement in reading comprehension. The purpose of this study was also to examine ways in which the washback effect could modulate the selection of reading texts in the classroom, their difficulty level, and their effectiveness from the perspective of test takers. A total of 48 state and private university students studying English language teaching (ELT) at the MA level took part in this study. Participants were asked to provide answers to a researcher-developed questionnaire as well as to open-ended questions. The results of data analysis revealed that the entrance exam had a positive washback on learners' reading comprehension in a number of ways. Furthermore, it was found that textbooks with a moderate level of difficulty were perceived as more effective by test takers. Implications for policymakers and suggestions for further research are presented.

Keywords: washback, assessment, reading comprehension, text difficulty, entrance exam

Introduction

It has been estimated that as much as half of teachers' time in classrooms is spent on assessment-related practices, (Stiggins, 1991) including designing, developing, selecting, administering, scoring, recording, reporting, evaluating, and revising assessment methods. The effectiveness of these assessment-related activities depends on learners' learning goals, which need to be scaffolded by professional support in order to face different challenging assessments necessary in the classroom.

The effect of assessment on teachers' practices in the classroom is closely related to the significant role that assessment plays for educational purposes. It is widely believed that effective assessment can improve students' achievement (Campbell & Collins, 2007); therefore, assessment plays an important role in teaching and learning, which can have effects on instructional decisions for teachers and administrators.

There are several systematic studies investigating the impact of Iranian entrance examinations, mostly focused on instructional approaches and syllabus design (e.g., Ghorbani, Samad, Hamzah, & Noordin, 2008; Ramezany, 2014; Rezvani & Sayyadi, 2016). To the researchers' knowledge, however, there have been no studies investigating how these tests are influencing teachers' selection of reading texts, with regards to their difficulty and developing learners' comprehension. Due to the dearth of research in this field and the value attributed to higher education by Iranian society, the present study aimed to look into the effect of the entrance examination in the context of the Master's examination course. For this purpose, this study employed an empirical research design that attempted to expand the current theories and body of research in the field of language testing and assessment.

Review of Literature

The importance and impact of testing and assessment are increasing around the world (Biggs, 1995). Public examinations

have undergone changes to meet teaching and learning needs, which can be seen in Australia, New Zealand, and countries in Europe and North America. The purposes behind these changes were to establish new national education systems or to modify current systems; moreover, policymakers attempted to design new assessment elements to promote teaching and learning and to provide accountability. As there is a growing interest in using language tests and scores in different contexts such as primary and higher education, employment, health and safety, industry, migration, citizenship policy and practice, language testing has become a big business (Spolsky, 1994). Accordingly, Shohamy (2007) discussed the highlighted role of language tests, their power, and their central function in education, politics, and society. Despite the increasing number of test takers, selectors, and developers using test scores for the purpose of decision-making, there is a seeming lack of background assessment or training (MacLellan, 2004). The center of attention is on classroom teachers' tasks that are based on standardized tests to evaluate students' improvement and to keep the school and teachers accountable for students' progress.

Since washback studies are by nature evaluation studies, they demand a comparison between what is on the ground and what should ideally be the case. The latter would push us to be utterly explicit about the ideal method of improving literary competence. Washback inquiries target the discrepancies between the real and the ideal and suggest ways to reduce such discrepancies. When evidence suggests that some or all of the identified discrepancies are in fact due to construct-irrelevance or construct-underrepresentation (Messick, 1996) in the test, findings from washback studies can be fed into test retrofits aimed at fostering further test validity. Given the complex nature of washback, as it is discussed in the literature, the current study attempted to shed further light on the individual and material factors in the process of washback, and to discuss the results in light of existing theories in applied linguistics.

2.1 Theoretical Framework

Cognitive Constructivist Theory

Constructivism is based on the assumption that students' and teachers' understanding and knowledge of the world can be constructed by themselves by having experiences and reflecting on those experiences (Vygotsky, 1978). Vygotsky defines the term "Zone of Proximal Development," or ZPD, as the distance between what the individual is able to do independently and what s/he will be able to do with the help of more capable other. In other words, the distance between the individual's actual and potential ability level is called the ZPD. Applied to the context of teaching, this means that teachers are all creative constructors of self-knowledge, which requires asking questions, exploration, and assessment of students' understanding. In a classroom context, this implies that testing, which is an ongoing process, needs to measure students' learning outcomes mostly in the form of formative testing rather than testing one. In other words, teachers get feedback from their students' improvements in learning and then provide them with feedback based on the results of their performance.

Brookhart (2004) and Cowie (2005) argue for the impact of testing on students' social construction as competent ones. Furthermore, students can be constructors of knowledge based on the constructivist theory of learning; therefore, the role of feedback as a major factor in the acquisition of knowledge is undeniable. Moreover, Andrews (2004) noted that the construction of meanings and beliefs by teachers and students in an educational, practice-based context is significant to theory and assessment practices modeled on the cognitive constructivist perspective. It needs to be noted that this theory forms the basis of the present study, in the sense that teachers' awareness and knowledge about testing in general, and how best to fit testing goals to their learners' needs, assist them in using testing for the purpose of learning. More importantly, this reflection about the integration of testing and learning informs teachers about the adoption of specific teaching activities and materials in order to achieve their preferred outcomes.

Sociocultural Theory

The present study is based on the theme that testing, even in the case of high-stakes tests, needs to be considered a dynamic process, that is, a part of a process of teaching and learning improvement that can help learners upgrade their knowledge. This viewpoint then fosters positive washback on both teachers and learners and is supported by the humanistic and social approach to language learning and teaching. Learning is a cognitive development through social interaction (Vygotsky, 1978), where the social context centers on the interaction between teachers' guidance, peers' collaboration, and learners' development in the Zone of Proximal Development. This theory declares that learners' potential to achieve higher levels of understanding is possible with adults' support, reactions, and interpretation, as well as interactions with capable peers, all of which should be in line with learners' cognitive development.

The sociocultural theory emphasizes that an individual's mental development can be achieved through meaningful verbal interactions with others in social contexts which involve complex and higher mental functions (Lantolf & Thorne, 2006). Vygotsky suggested that children move through two stages of development,

from potential development to actual development. The intermediary stage is named the Zone of Proximal Development (ZPD). Based on Vygotsky's ZPD theory (1978), learning and development are interwoven with each other. That is, learners need guidance from adults or peers, mediated by language, through social activities in order to achieve their learning goals. Sociocultural theory suggests that people use language, cultural artifacts, and other symbolic systems as mediation tools for their activities in these social contexts. This theory is based on the dynamic interdependence of social and individual processes and how humans use language to mediate activities, to construct knowledge, and to achieve their goals. Many pedagogical concepts (such as activity theory, scaffolding, the zone of proximal development, and dynamic assessment) are derived from this theory and have been applied to foreign and second language learning.

In his socio-cultural theory, Vygotsky (1978) proposed that humans use language to communicate with other people to share experiences and to construct knowledge in a society. He pointed out that the developing individual needs assistance with higher mental functioning development, which can be gained from other people's experiences through social interaction. That is, the mental development of an individual can be accomplished with assistance from other people in society through interaction. This viewpoint has influenced foreign and second language learning, both theoretically and practically.

Bailey's (1996) Washback Model

Compared to other washback models, Bailey's (1996) model is a simple and a comprehensive account of the distinct issues included in the washback process. This model comprises the participants, procedures, and outcomes involved in washback, and it attempts to delineate the ways a test can be connected with different stakeholders and how these stakeholders can impact one another.

Bailey's (1996) model of washback (Figure 1) actually stemmed from Hughes' (1993) distinction between participants, processes, and products, which illuminates the processes by which washback may operate. Put differently, he categorized the types of effects that might occur in the three components of participants, processes, and products in an educational system. "Participants" refer to classroom teachers and learners, educational administrators, material developers, and publishers. "Processes" are the practices conducted by participants that can bring about the process of learning, such as material development, syllabus design, creativity in teaching techniques, the implementation of test-taking strategies, and so forth. Finally, the term "products" refers to the learning objectives and the quality of learning (Hughes, 1993).

Based on Hughes' (1993) model, Bailey's model describes the direct impact of a test on several participants, all of whom are engaged in processes that lead to products specific to the participant. The dotted lines in Figure 1 display possible influences on the test from various participants. Bailey's (1996) model illustrates the straightforward influence that exposure to test-derived information has on test-takers, defining this as washback to the learners. She also describes the outcomes of test-derived information presented to teachers, administrators, curriculum designers, and therapists as washback to the program, based on Alderson and Wall's (1993) hypotheses.

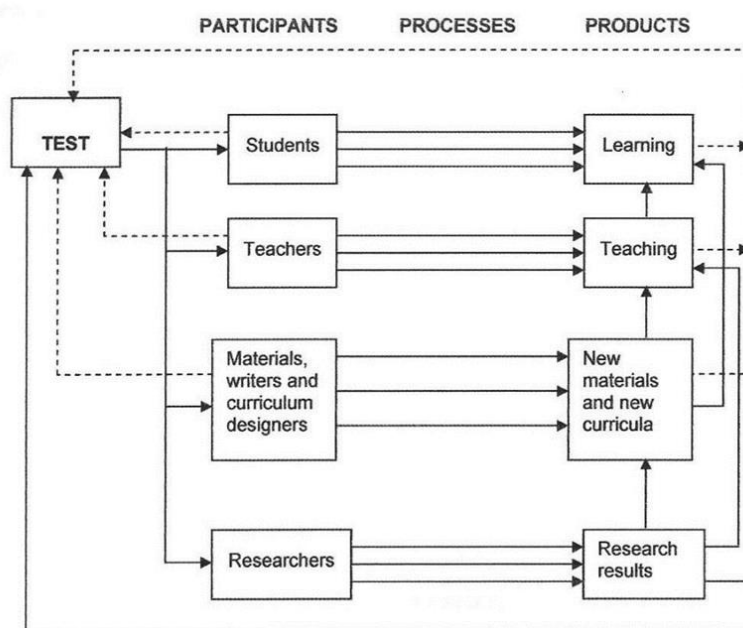


Figure 1. Washback model (Bailey, 1996)

2.2 Text Difficulty

One essential issue with regard to language learners' successful reading comprehension is the appropriate difficulty of the text. Weaver and Bryant (1995) proposed the Ideal Effort Theory, claiming that learners were more prepared to anticipate comprehension when content materials coordinated their reading level, instead of being too simple or excessively difficult. In other words, it is theorized that readers' capacity to foresee cognizance relies to a great extent on their reading level with respect to the difficulty of texts. Therefore, readers should have the capacity to anticipate their comprehension most precisely when texts' difficulties matches their reading level.

Reading instructors agree that pairing learners with materials of a suitable difficulty level is a standout amongst the most imperative choices they make. Before an instructor starts to teach any comprehension skill or procedure, the teacher must consider learners' experience with respect to the material being utilized and their reading level, and must also identify the purposes behind teaching. Cooper (1986) clarified that the text utilized for the introductory instruction of any comprehension skill or procedure should be at a difficulty level that is simple for the learner. If learners are intended to experience challenges in learning, the material should be at a personally challenging reading level, and if they are learning at an ordinary pace, the text's difficulty level can be at their instructional level. For learners who have trouble developing strategic behaviors, it is the teacher's instructional obligation to structure tasks that are sufficiently difficult, in order to oblige students to use strategic behavior, but simple enough that learners are able to finish the task with some effort (Garner, 1987). Since little research has been conducted concerning comprehension monitoring in terms of text difficulty, the inquiry then is concerned with what is too simple or what is too difficult.

Paris and Myers (1981) evaluated comprehension monitoring utilizing several distinct measures. The error detection task uncovered a low rate of error detection by both advanced and poor readers; on the other hand, when the results of different measures of comprehension monitoring were combined and

analyzed, an intriguing pattern was noted. Both advanced and poor readers did screen their comprehension; however, poor readers were not as precise and did not screen consistently. Poor readers saw a larger number of odd words than the advanced readers on the third-grade story, which was evidently simple for them; however, it was impossible to tell if these readers engaged in unsupervised monitoring of incomprehensible text on the fifth-grade story. This implies that poor readers do screen their understanding, yet how much and how precisely they screen may depend on the difficulty of the material. Different studies have proposed that comprehension is influenced by the type and difficulty of data in a text (Hare, 1981) and that learners are not taught to read different sorts of content in different ways (Raphael, Myers, Tirre, Fritz & Freebody, 1981).

2.3 Washback Studies

Qi (2005) explored the reasons why modifications to the format of the NMET (National Matriculation English Test) in China were not successful in leading to desired changes. The test developers' purpose was to encourage teaching and learning for authentic language use by adding listening multiple-choice items and a proofreading section and decreasing the number of grammar and vocabulary items. The results of the study by Qi (2005) indicated that, like many large-scale standardized tests, multiple-choice formats are not conducive to the development of effective teaching and learning activities that enhance real-life language use (see also Hamp-Lyons, 1998). Besides, to accomplish positive washback, the literature recommends that authentic materials and tasks be implemented (Bailey, 1996; Messick, 1996). Even though the writing and proofreading components of the test aimed to reflect a higher level of language in actual use, as Qi notes, in order to decrease subjectivity, the test content did not imitate real-life activities.

Nazari and Nikoopour (2011) surveyed the washback impacts of high school tests in 120 female Iranian high school learners' opinions on language learning. The researchers found that first, the participants concurred on the nature of the washback effect of the exams and, second, that there is a similarity between the

different effects of learners' language learning beliefs and their foreign language-learning processes. Likewise, Nikoopour, Amini Farsani and Nasiri (2012) investigated the washback influence of State and Azad UEE on Iranian EFL applicants. They discovered that UEE had an effect on students', learning strategies, and effective scope.

It can be seen that the distinctive feature of the positive impact of teaching in the form of washback is its foundation on a theory of learning and cognitive development. Guided by the socio-cultural theory of learning, positive washback distinguishes itself from negative effects in terms of how it is administered; namely, during the teaching practice, there is mediation/intervention (in a dynamic way) between the teacher and the learners to facilitate students' enhanced achievement.

Elder and O'Loughlin (2003) in their study concluded that contextual factors such as the types of accommodation that learners had lived in, course level, educational background, and reading capabilities together offered the best predictions of success in overall IELTS performance. Research by Clapham (1996) into the role of background knowledge in reading comprehension led to the conclusion that even though topic knowledge could be viewed as a potential contributor to comprehension in some contexts, there was no clear correspondence between general discipline fields and the kinds of background knowledge that could be evaluated.

In sum, according to the literature review, although there have been studies looking into the effects of high stakes examinations and their washback effects on learners' language development, including reading, there has been no research examining the effects on the difficulty of texts used in the classrooms. The present study focuses on a group of individuals admitted to the M.A Entrance Exam at one public university in Iran. In contrast to the investigation of washback within a specific classroom context, this study traces the motivating elements behind the preparation process of the test takers; examines the types of activities and processes that learners become involved in, and investigates major influences behind this. In addition to these factors, the novelty of this study results from the hypothesis that English majors are more vulnerable to washback as a result of their major (Shih, 2007), an issue that is the focus of the present study and a target for further elaboration and examination. This study seeks to answer the following research questions:

1. What is the washback effect of the English Language Teaching (ELT) MA entrance exam on learners' perceptions of the development of reading comprehension?
2. What is the washback effect of the English Language Teaching (ELT) MA entrance exam on the difficulty of texts used in the classroom?

Method

Participants

The study was carried out with 48 participants from three public universities in Iran: Urmia University, Tabriz University, and Islamic Azad University Tabriz Branch. None of the participants had spent any years in an English-speaking country. The ages of the participants ranged from 23-30. The mean age at which they started learning English was 9 years old. The average number of years in English instruction was fourteen years. Participants included both male ($N = 26$) and female ($N = 22$) learners. All participants were Masters students with majors in English language teaching. All participants orally consented to provide data for the present study.

Table 2. Confirmatory Factor Analysis Results

Instrument

Washback Effect Questionnaire

A washback effect questionnaire was developed by the researcher, based on research by Popham (2010) and Green (2007). The questionnaire consisted of two parts: Likert-type items and open-ended questions. The items were organized according to the factors of the entrance exam in Iran and the reading comprehension practices conducted to prepare learners for the exam. The questionnaire inquired about the practices carried out in the classroom in order to get insights into learners' perceptions about the adequacy of reading comprehension practices and the resulting output. The questionnaire was a Likert-scale inventory with a total of 15 items ranging from strongly disagree (1) to strongly agree (6). The reliability of the questionnaire was assured through a Cronbach's alpha, which showed a satisfying level of consistency ($\alpha = .90$, see Table 1 below for details). Furthermore, the questionnaire was found to be a valid indicator of the construct it was designed to measure, reflected by the results of Varimax rotation analysis, which showed that the items in the questionnaire were loaded with a 32.24% variance (see Table 2 for details). The questionnaire is provided in the Appendix.

In order to obtain information about the difficulty of the texts used in the class, learners were asked about the sources used in the classroom and their effectiveness. The reported sources were then analyzed for difficulty and appropriateness to the exam.

Procedure

The present study took the form of a survey, which attempted to examine washback effects of the MA entrance exam on ELT learners' reading comprehension and the difficulty of texts used in the classroom to prepare for the exam. Accordingly, the questionnaire was administered, either electronically or in person, to 48 Iranian English language teaching students in different universities. Completion of the questionnaire took no more than 10 minutes. The return rate of the questionnaires was 85 percent and information from those missing questionnaires was withdrawn.

Data Analysis

The results obtained from the questionnaire were analyzed with the Statistical Package for the Social Sciences (SPSS) version 21. For the first research question, descriptive statistics representing the means and standard deviations were calculated. For the second research question, the results of open-ended answers were analyzed in terms of frequency. The justification for the text difficulty was interpreted using existing criteria in language testing.

Results

Firstly, the questionnaire was subject to reliability and validity analyses to ensure its psychometric properties. The reliability analysis was carried out using Cronbach's alpha and the validity analysis was performed through confirmatory Factor analysis. The results are reported in Tables 1 and 2.

Table 1. Reliability Results

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.904	.906	15

Table 1. Confirmatory Factor Analysis Results

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4.837	32.247	32.247	4.837	32.247	32.247	3.015	20.098	20.098
2	2.189	14.591	46.837						
3	1.478	9.855	56.693						
4	1.239	8.258	64.951						
5	1.025	6.833	71.784						
6	.733	4.885	76.669						
7	.694	4.630	81.298						
8	.643	4.286	85.584						
9	.551	3.676	89.261						
10	.390	2.599	91.860						
11	.379	2.528	94.388						
12	.312	2.082	96.471						
13	.229	1.527	97.997						
14	.173	1.150	99.147						
15	.128	.853	100.000						

Extraction Method: Principal Component Analysis.

After the varimax rotation, a one-factor solution was chosen, which accounted for 32.24% of the total variance. These items

met the criterion of loading above 1.0 on their related factor. The descriptive statistics for each factor are displayed in Table 3.

Table 3. Descriptive Statistics for Each Questionnaire Item

Items	Mean	Standard Deviation	Minimum	Maximum
1. I learned vocabulary for my entrance exam in my last year BA courses.	3.90	1.40	1.00	6.00
2. I learned general vocabulary that was useful in the reading comprehension part of the exam.	4.15	1.38	1.00	6.00
3. I learned the about the types of tasks test takers are required to provide answers to in the entrance exam.	4.33	1.59	1.00	6.00
4. I learned ways of improving my English language reading comprehension test score.	4.30	1.54	1.00	6.00
5. I learned how to organize my time for studying.	3.90	1.57	1.00	6.00
6. My teacher corrected my mistakes in the classroom.	3.50	1.35	1.00	6.00
7. We worked on materials closely related to my future exam.	3.48	1.59	1.00	6.00
8. The activities we did in the class were similar to the ones on the entrance exam.	4.53	1.70	1.00	6.00
9. I learned quick and effective ways of reading books in English.	4.27	1.65	1.00	6.00
10. I read books and articles about my specialist subject area.	4.03	1.50	1.00	6.00
11. I learn how to arrange my time to read the exam texts.	3.94	1.65	1.00	6.00
12. I was able to take practice tests in class.	4.15	1.62	1.00	6.00
13. I learned useful learning strategies to help me prepare for the entrance exam.	3.13	1.69	1.00	6.00
14. I had a motivating course in order to better achieve a satisfying result in the exam.	3.23	1.54	1.00	6.00
15. My teacher clearly understood the objectives of this course.	3.35	1.58	1.00	6.00

According to participants' responses, the items with the highest mean scores reflected positive washback effect on the instructional activities done in the classroom and learners' perceptions of the development of their reading comprehension. For example, item 8 had the highest mean score ($M = 4.53$, $SD = 1.70$), followed by the third item ($M = 4.33$, $SD = 1.59$) and then

the fourth item ($M = 4.30$, $SD = 1.54$), all showing learners' beliefs about the test's positive effect on their reading comprehension.

The analysis of respondents' reported sources led to the identification of the following textbooks with the highest and lowest frequency of usage among the learners.

Table 4. Descriptive Statistics for Resources

			Fast Reading	Simple Prose Texts	TOEFL	GRE	Mahan	Pouran Pazhouhesh	Total
Sources	usage	Count	34	18	13	9	11	15	100
		% within items	34.0%	18.0%	13.0%	9.0%	11.0%	15.0%	100.0%
		% within choices	4.4%	2.6%	3.4%	2.0%	2.1%	2.3%	2.9%
		% of Total	1.0%	.5%	.4%	.3%	.3%	.4%	2.9%
	difficulty	Count	17	15	24	22	3	19	100
		% within items	17.0%	15.0%	24.0%	22.0%	3.0%	19.0%	100.0%
		% within choices	.32%	2.2%	3.1%	4.8%	.7%	2.9%	2.9%
		% of Total	.5%	.4%	.7%	.6%	.1%	.5%	2.9%
	efficacy	Count	25	20	1	12	16	26	100
		% within items	25.0%	20.0%	1.0%	12.0%	16.0%	26.0%	100.0%
		% within choices	3.2%	4.4%	.3%	1.7%	3.0%	3.9%	2.9%
		% of Total	.7%	.6%	.0%	.3%	.5%	.7%	2.9%

The results of frequencies and percentages show that the book with the highest use was Fast Reading, (34%) followed by Simple Prose Texts (18%) and Pouran Pazhouhesh (15%). The TOEFL and GRE books were largely not used in the classes, which can be attributed to the fact that the texts' difficulty was beyond learners' levels. As can be seen from Table 4, participants identified TOEFL (24%) and GRE (22%) as the most difficult texts, while Fast Reading (17%), Simple Prose Texts (15%) and Pouran Pazhouhesh (19%) had moderate levels of difficulty. Furthermore, the Mahan textbook (3%) was identified as the easiest source. Regarding effectiveness, respondents specified Pouran Pazhouhesh (26%), Fast Reading (25%) and Simple Prose Texts (20%) as the most effective resources that they used in the classes.

Discussion

The findings of this research revealed that MA students found the use of certain classroom activities, such as test-taking strategies, time organization, learning grammar and vocabulary, and so forth, to be effective in promoting their reading comprehension skills. This verified that assessment alteration can lead to intended positive washback, as was also described by scholars such as Pearson (1988) and Popham (1987). Accordingly, previous research also highlighted the significance of using authentic materials and tasks in the classroom to help learners develop appropriate strategies for knowledge development (Bailey, 1996; Khezrlou, 2012a, 2012b; Messick, 1996). This was also noted in an Iranian context, where Nikoopour, Amini Farsani and Nasiri (2012) showed the effect of the entrance examination on the way students prepare for the exam, their strategies, and their emotions. Therefore, in line with previous studies, the present study also reported that students' entrance exam preparation helped them determine what to learn and how to be prepared for the exam. This means that the use of exam preparation materials in the class could provide an authentic context for students, and could help them make sense of exactly what they could expect to encounter on the day of the exam. This finding, however, is in contrast with Qi's (2015) study, which found that exam preparation materials, such as multiple-choice test packages, were ineffective. These findings recommended that standards-based assessment led to changes in the ways that teachers taught, as well as in the content that they emphasized in the classroom. The findings of this study are in contrast with those reported by Alderson and Wall (1993) and Cheng (1997, 1998), who reported that high stakes tests usually modified what teachers instructed, but were not successful in changing how teachers taught. Entrance exams were also predicted to encourage learners to attempt higher levels of achievement (Cheng, 1998). This was supported by the findings of the current research, which indicated that entrance

exams encouraged some learners to try harder to gain better scores.

The findings also showed that the text factor was an important variable in determining learners' success. More specifically, books with a moderate level of difficulties such as Fast Reading, Simple Prose Texts, and Pouran Pazhouhesh, were the most frequently used sources and were also, in the opinions of students, more effective. The effect of text resources has been considered a particularly significant dimension of some washback studies (Alderson, 2004; Cheng & Curtis, 2004). The results of the previous research indicated that washback is influenced by numerous contextual factors and beliefs. This supported the argument of many scholars that the impact of washback was likely to be mediated by several elements; not just the test itself, but also contextual factors (Alderson, 2004; Brown, 2009; Shohamy, 1993; Wall, 1997; Watanabe, 1996). This clarifies why policymakers have often been unsuccessful in creating washback on teaching and learning simply by altering tests (e.g., Cheng, 1997, 1998; McNamara, 2000; Qi, 2005, 2007). According to Andrews (2004) and Watanabe (1996), washback, in fact, appears to be a highly complicated issue, as intended educational modifications can be advanced or hindered by different factors. Among these factors is textbook choice, as was the case in the present study.

Based on these findings, it can be concluded that the entrance exam was operating as assessment for learning, in the case of the reading comprehension part. In the past, washback has been considered unpredictable, due to the complexity of the processes behind it (e.g., Hayes & Read, 2004). The present study, in contrast, proposed that at least part of the mechanism of washback could be predicted, as the research exhibited ways in which particular washback effects occurred due to specific kinds of contextual factors.

Conclusion

The present study has significant implications for policymakers engaged in educational reform by modifying assessments. The generally held view is that intended washback could be caused only by proposing a new or altered assessment (e.g., Chapman & Snyder, 2000). This research, supporting the argument of many scholars in the areas of education and applied linguistics (e.g., Alderson, 2004; Chapman & Snyder, 2000; Fullan, 2001; Watanabe, 2004), highlighted the complicated nature of washback, due to the fact that it seems to be mediated by certain specific factors, such as the textbook resources used for test preparation. In order to be more effective, test preparation standards need to meet these two criteria by Popham (1993): 1) Professional ethics: as means of test preparation practices,

books should not violate the standards of education; rather, they should provide a solid base for test success, and 2) educational defensibility: all the suggested textbooks should enhance learners' knowledge of the subject as well as their test performance. The appropriateness and ethicality of the materials are also emphasized by Mehrens and Kaminski (1989, p. 16), who conclude that materials must adhere to the following criteria:

1. General instruction on objectives not determined by looking at the objectives measured on standardized tests;
2. Teaching test-taking skills;
3. Instruction on objectives generated by a commercial organization where the objectives may have been determined by looking at objectives measured by a variety of standardized tests. (The objectives taught may, or may not, contain objectives on teaching test-taking skills.);
4. Instruction based on objectives (skills, subskills) that specifically match those on the standardized test to be administered;
5. Instruction on specifically matched objectives (skills, subskills) where the practice instruction follows the same format as the test questions;

It is obvious that the intended washback cannot be implemented by modifying a test without considering other important factors, such as the ones mentioned above. It is predicted that fostering positive washback is possible when the opinions of the teacher and learner converge with the purpose of the Ministry of Education, and where contexts are supportive. By clarifying the relationship between assessment and preferred outcomes, this model can potentially help advance positive washback while diminishing unwanted negative washback.

This study has expanded understanding of the process of washback. Future research, however, can replicate this study in particular aspects. The questionnaire that was used in this research led to the analysis of learners' perceptions of washback. Therefore, the washback effects examined did not take into account what was actually taking place in the classrooms. Classroom observations, which are often applied by washback scholars in the area of applied linguistics, were not integrated into the current study because the attention was on investigating the role of contextual factors in washback. The use of classroom observations in future research could contribute to a richer image of washback than one based on participants' responses, and could well be the subject of future investigations to link washback to actual classroom behavior. Furthermore, in order to establish generalizability, it would be appropriate to have a larger sample size, particularly of learners in lower socioeconomic educational contexts. Further research is also needed to examine teachers' beliefs about the effect of the entrance exam upon their pedagogical activities and students' learning outcomes.

References

- Alderson, J. C. (2004). Forward. In L. Cheng, Y. Watanabe & A. Curtis (Eds.), *Washback in language testing: Research context and methods* (pp. ix-xii). Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.
- Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14(2), 115-129.
- Andrews, S. (2004). Washback and curriculum innovation. In L. Cheng, & Y. Watanabe with A. Curtis (Eds.), *Washback in language testing* (pp. 37-50). Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.
- Bailey, K. M. (1996). Working for washback: A review of the washback concept in language testing. *Language Testing*, 13(3), 257-279.
- Biggs, J. (1995). Assumptions underlying new approaches to educational assessment: Implications for Hong Kong. *Curriculum Forum*, 4(2), 1-22.
- Brookhart, S. M. (2004). Successful students' formative and summative uses of assessment. *Assessment in Education*, 8(2), 153-169.
- Brown, A. V. (2009). Students' and teachers' perceptions of effective foreign language teaching: A comparison of ideals. *The Modern Language Journal*, 93, 46-60
- Campbell, J., & Collins, D. (2007). *Student involved classroom assessment (3rd ed.)*. Columbus: Merrill Publishing.
- Chapman, D. W., & Snyder, C. W. (2000). Can high stakes national testing improve instruction: reexamining conventional wisdom. *International Journal of Educational Development*, 20(6), 457-479.
- Cheng, L. (1997). How does washback influence teaching? Implications for Hong Kong. *Language and Education*, 11(1), 38-54.
- Cheng, L. (1998). Impact of a public English examination change on students' perceptions and attitudes toward their English learning. *Studies in Educational Evaluation*, 24(3), 279-301.
- Cheng, L., & Curtis, A. (2004). *Washback or backwash: A review of the impact of testing on teaching and learning*. In L. Cheng, & Y. Watanabe with A. Curtis (Eds.), *Washback in language testing* (pp. 3-18). Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.
- Cooper, D. J. (1986). *Improving reading comprehension*. Boston: Houghton Mifflin.
- Cowie, B. (2005). Pupil commentary on assessment for learning. *The Curriculum Journal*, 16, 137-151.
- Elder, C., & Wigglesworth, G. (1996). Perspectives on the testing cycle: Setting the scene. *Australian Review of Applied Linguistics, Series S*(13), 1-12.
- Fullan, M. (2001). *The new meaning of educational change (3rd ed.)*. New York: Teachers College,
- Columbia University. Garner, R. (1987). *Metacognition and reading comprehension*. Norwood, NJ: Ablex.
- Ghorbani, M. R., Samad, A. A., Hamzah, M. S. G., & Nooreen, N. (2008). The washback effect of the Iranian university entrance examination on pre-University English teachers. *The Journal of Asia TEFL*, 5(3), 1-29.
- Green, A. (2007). *IELTS washback in context: Preparation for academic writing in higher education*. Studies in language testing (Volume 25). Cambridge: Cambridge University Press.
- Hamp-Lyons, L. (1998). Ethical test preparation practice: The case of TOEFL. *TESOL Quarterly*, 32, 329-337.

- Hare, V. C. (1981). Readers' problem identification and problem-solving strategies for high and low knowledge comprehenders. *Journal of Reading Behavior*, 13, 359-365.
- Hayes, B., & Read, J. (2004). IELTS Test Preparation in New Zealand: Preparing Students for the IELTS Academic Module. In L. Cheng, Y. Watanabe & A. Curtis (Eds.), *Washback in Language Testing* (pp. 97-111). Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.
- Hughes, A. (1993). *Backwash and TOEFL 2000*. Reading, U.K.: University of Reading.
- Khezrlou, S. (2012a). The relationship between cognitive and metacognitive strategies, age and level of education. *The Reading Matrix*, 12(1), 50-61.
- Khezrlou, S. (2012b). Cognitive strategy training: Improving reading comprehension in the language classroom. *Journal of Teaching Language Skills*, 3(4), 77-98.
- Lantolf, J. P., & Thorne, S. L. (2006). *Sociocultural theory and the genesis of second language development*. Oxford: Oxford University Press.
- MacLellan, E. (2004). Initial knowledge states about assessment: Novice teachers' conceptualizations. *Teaching and Teacher Education*, 20, 523-535.
- McNamara, T. (2000). *Language testing*. Oxford: Oxford University Press.
- Mehrens, W. A., & Kaminski, J. (1989). Methods for improving standardized test scores: Fruitful, fruitless, or fraudulent. *Educational Measurement: Issues and Practices*, 8(1), 14-22.
- Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13, 241-256.
- Nazari, M., & Nikoopour, J. (2011). The wash back effect of high school examinations on Iranian EFL learners' language learning beliefs. *Journal of Language Translation*, 2(1), 29-49.
- Nikoopour, J., Amini Farsani, M., & Nasiri, M. (2011). On the relationship between critical thinking and language learning strategies among Iranian EFL learners. *Journal of Technology & Education*, 5(3), 195-200.
- Paris, S. G., & Myers, M. II. (1981). Comprehension monitoring memory and study strategies of good and poor readers. *Journal of Reading Behavior*, 13(1), 5-22.
- Pearson, I. (1988). *Tests as levers for change*. In D. Chamberlain & R. J. Baumgardner (Eds.), *ESP in the classroom: Practice and evaluation* (pp. 98-107). London: Modern English.
- Popham, W. J. (1987). The merits of measurement-driven instruction. *Phi Delta Kappa*, 68, 679-682.
- Popham, W. J. (1993). Measurement-driven instruction as a "quick-fix" reform strategy. *Measurement and Evaluation in Counseling and Development*, 26(1), 31-34.
- Popham, W. J. (2010). *Demystifying the tools that shape our schools*. Pearson.
- Qi, L. (2005). Stakeholders' conflicting aims undermine the washback function of a high-stakes test. *Language Testing*, 22(2), 142-173.
- Qi, L. (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: Principles, Policy & Practice*, 14(1), 51-74.
- Ramezane, M. (2014). The washback effect of university entrance exam on Iranian EFL teachers' curricular planning and instruction techniques. *Procedia-Social and Behavioral Sciences*, 98, 1508-1517.
- Raphael, T. E., Myers, A. C., Tirre, W. C., Fritz, M., & Freebody, P. (1981). The effects of some known sources of difficulty on meta-comprehension and comprehension. *Journal of Reading Behavior*, 13, 324-334.
- Rezvani, R., & Sayyadi, A. (2016). Washback effects of the new Iranian TEFL Ph.D. program entrance exam on EFL instructors' teaching methodology, class assessment, and syllabus design. *Theory and Practice in Language Studies*, 6(5), 1111, 2016.
- Shih, C. M. (2007). A new washback model of students' learning. *The Canadian Modern Language Review*, 64(1), 135-162.
- Shohamy, E. (1993). *The power of test: The impact of language testing on teaching and learning*. Washington, DC: National Foreign Language Center.
- Shohamy, E. (2007). The power of language tests, the power of the English language and the role of ELT. In J. Cummins & C. Davison (Eds.), *International handbook of language teaching* (pp. 521-532). New York: Springer.
- Spolsky, B. (1994). The examination-classroom backwash cycle: Some historical cases. In D. Nunan, R. Berry & V. Berry (Eds.), *Bringing about change in language education: Proceedings of the International Language in Education Conference 1994* (pp. 55-66). Hong Kong: University of Hong Kong.
- Stiggins, R. J. (1991). Assessment literacy. *Phi Delta Kappan*, 72(7), 534-539.
- Vygotsky, L. (1978). Interaction between learning and development. In B. Gauvain & F. Cole (Eds.) *Readings on the development of children* (pp. 34-40). New York: Scientific American Books.
- Wall, D. (1997). Impact and washback in language testing. In C. Clapman & D. Corson (Eds.), *Encyclopaedia of language and education: Vol. 7. Language testing and assessment* (pp. 291-302). Dordrecht: Kluwer Academic.
- Watanabe, Y. (1996). Does grammar translation come from the entrance examination? Preliminary findings from classroom-based research. *Language Testing*, 13(3), 318-333.
- Weaver III, C. A., & Bryant, D. S. (1995). Monitoring of comprehension: The role of text difficulty in metamemory for narrative and expository text. *Memory and Cognition*, 23(1), 12-22.