

The Effect of Washback on Reading Comprehension of Medical Students in English for Specific Purposes Classes

Mohammad Naser Shafiee Jafarabadi¹; Nazila Zarghi²; Vahideh Zolfaghari³; Mohammad Reza Kargozari^{1*}

¹ English Department, school of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran

² Department of Nursing, school of Nursing, Mashhad University of Medical Sciences, Mashhad, Iran

³ Department of Computer Sciences, Imam Reza University, Mashhad Iran

* English Department School of Medicine Azadi Square Mashhad, IRAN
Postal code: 91779-48564
Tel: +98 513 800 2174
Fax: +98 513 800 2173
Email: Kargozarmr1@mums.ac.ir
Received: August 4, 2014
Accepted: November 29, 2014

Background: Testing and teaching are interrelated in education process; using various kinds of tests, teachers are able to anticipate the strong and weak points of learning in students, their progress and their accomplishment. The influence of test on teaching and learning is commonly referred to as washback. Testing washback is a twisted concept that becomes even more complex under a various interpretations of the washback phenomenon on teaching and learning. This study aimed to describe the effects of washback on behaviors of students, in the low stakes testing in English for Specific Purpose (ESP) environment in two-fold.

Methods: The effects of different formative tests on the medical students' English reading comprehension were assessed using Michigan test. The effects of washback on the students' attitudes toward English reading comprehension were measured by the English Reading Attitudes Questionnaire (ERAQ). Data were analyzed by paired t-test, t-test, and Mann-Whitney U test.

Results: Formative tests did not significantly affect on students' reading comprehension achievement in an ESP environment with (Mean = 71.77 ± 12.3 for experimental group vs. 67.59 ± 10.3 for control group, p > 0.05). But they showed significant effects on students' English reading attitudes (Mean = 63.23 ± 7.6 for experimental group vs. 54.13 ± 8.7 for control group; p < 0.001).
Conclusions: Findings from this study indicate that although formative tests do not significantly affect students' reading comprehension achievement in English for Specific Purpose environment, they have significant effects on students' English reading attitudes.

Keywords: Washback; Formative Test; Reading Attitude; English for Specific Purpose.

تأثیر بازآورد آزمون بر خواندن و درک مطلب دانشجویان پزشکی در کلاسهای انگلیسی به منظور خاص

زمینه و هدف: در فرآیند آموزش، ارزشیابی و تدریس در هم تنیده شده اند. به طوری که با استفاده از انواع مختلف ارزشیابی معلمها قادرند نقاط ضعف و قوت دانش آموزان، پیشرفت و موفقیت آنها را در امر یادگیری پیش‌بینی نمایند. تأثیر آزمون بر تدریس و یادگیری را بازآورد می‌نامند. بازآورد آزمون مفهوم پیچیده‌ای دارد که با تفسیرهای مختلف پدیده بازآورد در تدریس و یادگیری، پیچیده تر نیز شده است. این مطالعه در نظر دارد تأثیرات بازآورد آزمون را در رفتار دانشجویان توصیف نماید. جامعه هدف در این آزمون‌ها محدود بوده و در محیط استفاده تخصصی از زبان انگلیسی در دو بعد اجرا شده است.

روش: تأثیر آزمونهای مستمر بر یادگیری خواندن و درک مطلب انگلیسی دانشجویان پزشکی با استفاده از آزمون میسینگان بررسی شد. تأثیر بازآورد آزمون بر دیدگاه دانشجویان نسبت به خواندن و درک مطلب انگلیسی به وسیله پرسشنامه نگرش خواندن انگلیسی (ERAQ) اندازه گیری شد. داده ها توسط آزمون میانگین زوجی، آزمون t، و آزمون مان ویتنی تحلیل شد.

یافته‌ها: آزمون‌های مستمر بر فراگیری خواندن و درک مطلب انگلیسی دانشجویان در محیط استفاده از زبان انگلیسی تخصصی پزشکی با (میانگین = ۷۱/۷۷ + ۱۲/۳) برای گروه هدف، در مقابل ۶۷/۵۹ + ۱۰/۳ برای گروه کنترل (P > 0.05) تأثیر معناداری نداشت. اما تأثیر معناداری را در نگرش دانشجویان در خواندن انگلیسی نشان می دهد. (میانگین = ۶۳/۲۳ برای گروه هدف، و ۵۴/۱۳ + ۸/۷ برای گروه کنترل و P = ۰/۰۰۱).

نتیجه گیری: یافته‌های این مطالعه حاکی از آن است که اگرچه آزمونهای مستمر بر فراگیری خواندن و درک مطلب انگلیسی دانشجویان در محیط استفاده از زبان انگلیسی تخصصی پزشکی به طور معناداری تأثیر گذار نیست، اما بر نگرش دانشجویان در خواندن انگلیسی تأثیر معناداری دارد.

کلیدواژه‌ها: بازآورد آزمون، آزمون مستمر، نگرش خواندن، زبان انگلیسی به منظور خاص

تأثیر نتیجه‌ی الاختبار على القراءة و استيعاب المطلب عند طلاب الطب في صفوف درس اللغة الانجليزية لاهداف خاصة

التمهيد و الهدف: هناك اندماج وثيق بين التقييم و التدريس في عملية التعليم. إن استعمال اساليب مختلفة في التقييم يجعل المدرسين قادرين على ايجاد نقاط الضعف و القوة عند الطلاب مما يتيح لهم الفرصة في اجراء برامج افضل في مجال التعلم. يطلو عبارة (اعادة العطاء) على تأثير الاختبار على التدريس و التعلم. إن عبارة اعادة العطاء مفروم مقدر و زاد من تعقيداتها التفاسير المتعددة في مجال التدريس و التعلم. إن الهدف من هذه الدراسة هو توصيف مستوى اعادة عطاء الاختبار على سلوك الطلاب. هناك عدد محدود من الطلاب تم اشراكهم في هذه الدراسة و تم الاجراء في محيط تخصصي في مجال اللغة الانجليزية في بعدين .

المنهجية: تم تحليل تأثير الاختبارات المتتالية على تعلم القراءة و استيعاب المطلب الإنجليزي با استخدام ميشيگان. تم استخدام استمارة (ERAQ) رويته قراءة الإنجليزية و هذه الاستمارة كانت تريف الى تعيين مستوى تأثير اعادة العطاء. من خلال روية الطلاب في مجال القراءة و استيعاب المطلب في اللغة الانجليزية. تم تحليل المعطيات بواسطة معايير احصائية منرا اختبار T و اختبار مان و ويتني و اختبار المعدل التناسلي.

النتائج: لم يكن هناك تأثير ذوقيه معنى سرا في مجال اجراء اختبارات متتالية على تعلم القراءة و استيعاب المطلب الإنجليزي في مكان استخدام اللغة الإنجليزية التخصصية الطبية (معدل ۷۱.۷۷ ± ۱۲.۳ لفرقة الهدف. في مقابل ۶۷.۵۹ ± ۱۰.۳ لفرقة التأكيد (P > ۰.۰۵)). و لكن هناك تأثير ذوقيه في روية الطلاب في مجال قراءة الإنجليزية (المعدل = ۶۳.۲۳ ± ۷.۶ لفرقة الهدف في مقابل ۵۴.۱۳ ± ۸.۷ لفرقة التأكيد (P = ۰.۰۰۱)).

الاستنتاج: اذتارت نتائج هذه الدراسة الى أن لم يكن هناك تأثير يذكر في مجال أخذ اختبارات متتالية في مجال تعلم القراءة و استيعاب المطلب الإنجليزي في اماكن استعمال اللغة الإنجليزية التخصصية الطبية و لكن هناك تأثير في روية الطلاب تجاه قراءة الإنجليزية.

الكلمات الرئيسية: اعادة عطاء الاختبار، الاختبار المتتالي، اللغة الإنجليزية لأغراض خاصة .

خاص اهداف کے تحت انگریزی تعلیم کا جائزہ

بیک گراؤنڈ: تعلیمی نظاموں میں تعلیم اور اس کی کیفیت کا جائزہ لینا ایک دوسرے کے ساتھ جڑے ہوئے ہوتے ہیں۔ تعلیمی کیفیت کا جائزہ مختلف طریقوں سے لیا جاتا ہے اس کے ذریعے معلم اپنے طلباء کی کمزوریوں اور توانائیوں سے آگاہ ہوجاتا ہے اور ان کی علمی پیشرفت کے بارے میں اندازہ لگاسکتا ہے۔

روش: تعلیم و تدریس کے عمل کا جائزہ لینے کی روش کو واش بیک کہا جاتا ہے۔ اس تحقیق میں میڈیکل طلباء کے پڑھنے لکھنے اور انگریزی پڑھنے کے عمل کا جائزہ میسینگان ٹسٹ سے لیا گیا۔ انگریزی کے تعلق سے طلباء کے نظریات کا جائزہ اور یہ کہ وہ کس حد تک انگریزی سمجھتے ہیں آئی آر اے کیو ٹسٹ سے لیا گیا۔ ڈیٹا کا تجزیہ ٹی ٹسٹ اور مان ویتنی ٹسٹ سے کیا گیا۔

نتیجے: مستقل طور سے طلباء کے ٹسٹ لینا اور یہ جانچنا کہ وہ کس حد تک انگریزی میں پیشرفت کرچکے ہیں اور وہ بھی میڈیکل کالجوں اور یونیورسٹیوں کے ماحول میں کسی خاص معنی کا حامل نہیں ہے۔

سفارشات: گرچہ اس تحقیق سے معلوم ہوا ہے کہ واش بیک سے طلباء میں انگریزی کی تعلیم حاصل کرنے کے محرکات پیدا نہیں ہوتے ہیں لیکن بہر حال طلباء کے نظریات پر اثر انداز ہوتے ہیں۔

کلیدی الفاظ: انگریزی کی تعلیم، نظریات، جائزہ۔

INTRODUCTION

Testing has penetrated in all aspects of our life. It also influences on all grades of education, i.e. elementary, secondary, high school and university. Wherever there is teaching, there is also testing. On the other hand, the relationship between teaching and testing has long been concerned with both educational and applied linguistics literature. The fact that tests attract classroom teaching and the syllabus and make the teachers teach for examinations, and tempt students to do the exercises which are necessary to pass the examination is usual. Tests are generally used to make, among other things, inferences about examinees' abilities, predictions about their future success, and decisions (such as employment, placement, selection, etc.) about the test-takers. The uses made of test results thus imply values and have an impact on, the society and the educational system in general, and the students in particular(1).

In applied linguistics research, this influence of testing on teaching and learning has been referred to as "washback". The following definitions of 'washback' represent its meaning as used in the paper. Buck(2) defined it as the influence of the test on the classroom. This washback effect can be either beneficial or harmful. Messick(3) declares the extent to which the test acts upon language teachers and learners to do things that they would not necessarily otherwise do, and Bailey(4) defined it as the influence of testing on teaching and learning.

The test becomes a part of educational process in which changes in the educational system take place based on feedback obtained from the test. Public examinations that concern to high stakes influence more learners and teachers than class examinations that concern to low stakes, and consequently public examinations will have more difficult statistical analysis and interpretation than class examinations that are limited to students of one or two classes. However, the results of low stakes examinations are more explicit; they represent the students' learning condition, their strong and weak points.

It has long been believed that tests directly affect the educational processes in different ways. One assumption is that teachers will be influenced by the knowledge that their students need to take a certain test and will adapt their instruction and lesson content to cover the test. By test washback teachers learn more about students learning and their own teaching. Washback has been defined differently by different researchers, based on different aspects they study. Some researchers focused on learner and learning; some studied its effects on teacher and teaching, material, curriculum, school and even educational organization and society.

Buck(2) in his definition emphasizes the impact of a test on what teachers and students do in classrooms. One of the most common definitions referred to as the influence of testing on teaching and learning (e.g., Alderson & Wall(5); Cheng & Curtis(6)). Similarly, Brown(7) refers to washback as the effect of test on teaching and learning. Shohamy(8), focuses on washback in terms of language learners as

test-takers when she explains the use of external language tests to influence on foreign language learning in the school context. In another article, Shohamy, Donitsa-Schmidt, and Ferman(9), define washback as the connections between testing and learning. Bailey(4) also takes a broad view. He describes washback in terms of how tests influence educational practices and beliefs. More recently, Bachman and Palmer(1) define washback as a subset of a test's impact on society, educational systems, and individuals. They mention that test impact operates at two different levels: the micro level (i.e., the effect of the test on students and teachers) and the macro level (the effect on society and its educational systems). Cheng(10) also conducted research in the context of secondary school exams in Hong Kong. She explains washback as "an active direction and function of intended curriculum change by means of the change of public examinations, as is the case under study in Hong Kong"(10). Valette in (11) stated that "washback occurs when it is the testing instrument rather than the statement of desired learner outcomes that determines the nature of the curriculum and the course of instruction"; and Saville in (12) defines it as the influence of testing on teaching and learning. Also Alderson and Wall(5) claimed that "the extent to which the test influences language teachers and learners to do things they would not otherwise necessarily do" (p. 117). Messick(3), noting that washback can have either harmful or positive effects, expands it as: We emphasize first the need to establish valid evidential grounds for trustworthy inferences about tested language proficiency to provide a basis for distinguishing test-linked positive washback from good teaching regardless of the quality of the test and negative washback from poor teaching. (3)

It is worth mentioning that in recent studies, washback encounters a broader definition, taking the view that tests can have more far-reaching influences within the field of education. For example, Andrews (13), explores the relationship between washback and curricular innovation, beyond the classroom, and depict the effects of tests on teaching, learning, the educational process, and different stakeholders in the educational system. He uses washback to refer to: "the effects of tests on teaching and learning, the educational system, and the various stakeholders in the education process" (13).

The general objective of the present study was to describe the washback effect of the English Language Tests on reading comprehension in an ESP environment for medical students. In fact, this study had three main objectives:

1. to investigate the effect of different formative tests on reading comprehension in English for Specific Purpose (ESP) environment;
2. to investigate the effect of washback on attitudes of learners; and
3. To suggest areas for further research on washback.

METHODS

Participants

The participants of this study were four classes consisting of 130 students in total, 64 students in experimental group and 66 students in control group that had enrolled in the

Nursing School of Mashhad University of Medical Sciences (MUMS). The reason for choosing them is that they did not have any background of English as specific purpose; and because this university is one of the five, type one Medical University in Iran. The participants' mother tongue was Persian, and all participants had been learning English as a foreign language at least for five years. The setting of the present study was a compulsory English course for all medical freshmen students in any major. The two classes were held three hours a week and lasted for 16 weeks. The course instructor was the researcher.

Sampling Procedure

There are 130 medical students who enrolled at Mashhad University of Medical Sciences and have English course1 in four different classes, which two classes were randomly chosen as experimental and two classes as control group. By administering the Michigan Test of English Proficiency, in the very first session, the homogeneity of two groups was approved.

Design and Procedures

In order to measure the effectiveness of different formative tests on reading comprehension in ESP environment the decision was taken to conduct a research experiment which as Dörnyei(14) alleged was considered as the special quantitative data collection design, "which many people would claim to represent quantitative research at its most scientific because it can establish unambiguous cause-effect relationships"(p.115). After administering the Michigan Test of English Proficiency and homogenizing two classes of students, two of the classes randomly was chosen as experimental class (with 30 and 34 students) and the other two classes as control class (consisting of 31 and 35 students). The material was taught to both groups with the same method, and both groups received pre-test and post-test. Of course only the experimental group received the treatments, which were different formative tests during the course. These tests meanwhile are the independent variables of this research; while students' ability in reading comprehension and their attitudes toward English reading is dependent variables in this study, so sex and prior knowledge are control variable of this study.

Instructional Procedure

The material was taught based on the method, which is prescribed by the Center for Studying and Compiling University Books in Humanities (SAMT). Since this traditional text-driven teaching method produces no English speakers but rather English readers, the researcher used Chastain(15) teaching reading strategies. In experimental classes after covering each lesson, students received a specific related formative test for that lesson. Therefore, in experimental group each student passed 8 tests during one semester, plus a pre-test and post-test, Michigan English language proficiency test. It is worth mentioning that all of these teacher-made exams were approved by English department at MUMS, and their reliabilities were calculated by SPSS software. In control classes the material and the way of presenting the course is similar but they only received two tests, Michigan test of English language proficiency as pre-test and post-test.

To know whether these formative tests had any effect on experimental students' attitudes, another instrument was used. The quantitative data of this part of the study was collected via the English Reading Attitude Questionnaire (ERAQ) developed by Hung and used by Chiang (12) a master thesis to explore participants' attitudes toward EFL reading. Students received this questionnaire one month after their post-test. It consisted of twenty-one items all were closed questions designed using four-point Likert scale for easier analysis of the results. Its points start from Strongly Agree to Strongly Disagree as followings:

SA=4, Strongly Agree; A=3, Agree; D=2, Disagree; SD=1, Strongly Disagree.

The questionnaire was anonymous in order to encourage honest responses and the students were informed of this fact before it was distributed, it was also mentioned that it has no effect on their scores. Informed consent was taken from participants. Data were analyzed by paired t-test, t-test, and Mann-Whitney U test.

The research project is registered in Clinical Trials.gov (Code NCT016423).

RESULTS

Participants aged from 18 to 21 years old, the average age was 19. The first section of study deals with the students' performance on pre- and post-tests. And the second section reports the findings of the students' attitudes towards reading in English as measured by the English Reading Attitudes Questionnaire (ERAQ).

Within Group Pre- vs. Post-Test Analysis

The experimental group' Michigan test score increased from 62.7 ± 9.39 (Mean \pm SD) in the pre-test to 71.76 ± 12.31 (Mean \pm SD) in the post-test ($t = -9.6$) and the critical value of t ($t_{crit} = 2.045$).

The control group took both the pre- and post-tests as the experimental group did. The experimental group' Michigan test score increased from 63.37 ± 9.15 (Mean \pm SD) in the pre-test to 67.58 ± 10.28 (Mean \pm SD) in the post-test ($t = -8.86$) and the critical value of t ($t_{crit} = 2.048$).

Between Groups Analysis

The two-tailed test of significance for pre-test scores by the experimental and control groups indicates that there was no significant difference in pre-test scores between experimental and control groups ($p > 0.05$). Therefore, it is assumed that at the beginning of the study two groups must have come from the same population (Mean Difference = -0.67, Std. Error Difference = 2.41, $t(57) = -0.281$, Sig. 2-tailed = 0.78, $p > 0.05$).

As for the *post-tests* there was no significant difference between experimental and control groups as well ($P > 0.05$). Therefore, it is assumed that improvement of reading comprehension in both experimental and control groups is similar. In other word, the post-tests in experimental and control groups must come from the same population because no significant differences exist (Mean Difference = 4.18, Std. Error Difference = 2.95, $t(57) = 1.41$, Sig. 2-tailed = 0.16 $p > 0.05$).

Questionnaire Findings

The results of twenty one items of English Reading Attitude

Questionnaire (ERAQ) in two groups are (Mean Difference = 9.09, Std. Error Difference = 2.11, $t(57) = 4.29$, Sig. 2-tailed = 0.0 $p < 0.05$).

For more confirmation Mann-Whitney U test was also conducted. A significant difference exists between English reading attitudes in two different groups according to Mann-Whitney U test ($z = -3.67$, Sig. 2-tailed = 0.0, p

DISCUSSION

Results of this study shows that there is a significant difference between pre-test and post-test scores obtained by the students in both experimental and control group. There was no significant difference between the experimental and control group at the beginning of the study.

There was no significant difference in post-test scores between experimental and control groups as well. The two-tailed test of significance for post-test in experimental group and control group is more than .05. To put it in other words, there was no significant difference between the two groups, and it means that washback of formative tests did not affect reading comprehension in experimental group significantly. Therefore, the first null hypothesis was accepted.

To find the influence of washback on attitudes towards reading comprehension in English, students in the control and experimental groups received a questionnaire at the end of the course. Analyzing the results of students' questionnaire scores display that with $t(57) = 4.296$; $p < .05$, there was significant difference between students' attitudes in two different groups. So the null hypothesis was rejected, and its alternative hypothesis was accepted.

In terms of pedagogical implications, the findings of this study suggest that, first, formative tests influence teachers' and learners' behavior, and help them to find their weak points, and to improve their understanding of unknown

parts of the course. These tests may not improve their reading comprehension significantly, but they have significant effects on students' attitudes. Second, formative tests affect on the way and the content of teaching. By knowing the weak areas of students' language knowledge, the focus of teaching will be changed to those areas, and the teacher spends more time on them.

In terms of research implications, investigation of the data reveals that it is better to extend the categories of study and analysis beyond the scope of tests in class. The researcher should take the advantage of using different teachers in his or her study, so the results will be more convenient.

Although large strides have been made in identifying factors that might influence washback since Alderson and Wall's (5) call for more empirical research, only recently have studies begun pinpointing instances that may have had a direct impact on the amount of washback observed in specific situations, which is necessary if generalizations are ever to be made. Therefore, more case studies identifying factors affecting washback need to be carried out. Most of the washback research is confined to teachers, but all teachers and researchers believe in the fact that learners are the main determination of any learning process.

Future research should gather evidence from different populations so that the phenomenon of washback can be understood more thoroughly.

Replication of the study on different subjects and populations in order to increase generalizability is also recommended.

ACKNOWLEDGEMENT

Research committee approval and financial support: This research was financially supported by Mashhad University of Medical Sciences. The research project is also registered in Clinical Trials.gov (Code NCT016423).

Conflict of Interest: Authors declare no conflict of interest

REFERENCES

- Bachman LF, Palmer AS. Language testing in practice: Designing and developing useful language tests. USA: Oxford University; 1996.
- Buck G. Testing listening comprehension in Japanese university entrance examinations. JALT Journal 1988; 10(1): 15-42.
- Messick S. Validity and washback in language testing. Language testing 1996; 13(3): 241-56.
- Bailey KM. Working for washback: A review of the washback concept in language testing. Language testing 1996; 13(3): 257-79.
- Alderson JC, Wall D. Does washback exist? Applied linguistics 1993; 14(2): 115-29.
- Cheng L, Curtis A. Washback or backwash: A review of the impact of testing on teaching and learning. In: Cheng L, Watanabe Y, Curtis A. Washback in language testing: Research contexts and methods. New Jersey: Lawrence Erlbaum ; 2004: 3-17.
- Brown JD. University entrance examinations: Strategies for creating positive washback on English language teaching in Japan. Shiken: JALT testing and evaluation SIG Newsletter 2000; 3(2): 4-8.
- Shohamy E. Beyond proficiency testing: A diagnostic feedback testing model for assessing foreign language learning. Modern language journal 1992; 76(4): 513-21.
- Shohamy E, Donitsa-Schmidt S, Ferman I. Test impact revisited: Washback effect over time. Language testing 1996; 13(3): 298-317.
- Cheng L. How does washback influence teaching? Implications for Hong Kong. Language and education. 1997; 11(1): 38-54.
- Chen LM. Washback of a public exam on English teaching. New York: Wiley; 2002.
- Chiang CI. The effects of graphic organizers on Taiwanese tertiary students' EFL reading comprehension and attitudes towards reading in English. Australia: Australian Catholic University; 2005.
- Andrews S. Washback and curriculum innovation. In: Cheng L, Watanabe Y, Curtis A. (editors). Washback in language testing: Research contexts and methods. Mahwah, NJ: Lawrence Erlbaum Associates; 2004: 37-50
- Dörnyei Z. Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies. Oxford: Oxford University; 2007.
- Chastain K. Developing second-language skills. 2nd ed. Theory to practice. Boston: Houghton Mifflin; 1976.