

High-stake Test Preparation Courses: Washback in Accountability Contexts

Sima Sadeghi¹

Abstract

Given the increasing public demands for preparation courses, the purpose of this study was to discover how high-stakes testing influenced preparation courses. To achieve this end, concepts from Structuration Theory and Washback Hypothesis were employed. The study was conducted with four instructors teaching at the preparation courses for TOEFL and IELTS in Iran. An interpretive ethnographic case study was conducted through observation and field notes to gather data about how high-stakes testing affected teachers' curriculum and methodology. In order to keep observation focused and directed, University of Cambridge Observation Scheme (UCOS) was also used. The scheme was utilized to solely focus on the extent to which the course was test-oriented. The results indicated that teachers were constantly encountered with challenging questions which led to variations in their responses to tests' pressures.

Keywords: High-stakes Testing; Preparation Courses; Washback; Classroom observation

1. Introduction

Testing, teaching, and learning constitute three intertwined domains of Applied Linguistics. For long, scholars have argued how measurement could act as a powerful leverage and curriculum magnet which brings about improvement in learning and teaching.

¹Assistant Professor, PhD in English Language Teaching, University of Hormozgan, Iran, English Department; Faculty of Humanities, University of Hormozgan, Bandar Abbas, Iran, Email: sima_sadeghi23i@yahoo.com

Measurement-driven Instruction (Popham, 1985); *Curriculum Alignment* (Shepard, 1991); *Systemic Validity* (Fredericksen & Collins, 1989); *Washback Validity* (Morrow, 1986); *Consequential Validity* (Messick, 1996) though distinct, are all related to the test effect. The effect of the test on teaching, learning and other educational and social context is referred to as *Impact* (Bachman & Palmer, 1996), *Washback* (Alderson and Wall, 1993) and *Backwash* (Biggs, 1995). This study focuses on washback as defined by Bailey (1996) "the influence of testing on teaching and learning in the classroom"(p, 259), which could range from methodology, materials, to participants' feeling and attitudes.

Researchers have approached the issue of washback from different perspectives: *existence* of washback (Wesdorp, 1982; Shohamy, 1993), *direction* of washback (Frederiksen, 1984; Stake, 1991), *theoretical models* of washback(Alderson and Wall, 1993; Hughes, 1993, Bailey, 1996) and *scopes* of washback (Alderson and Wall, 1993; Bailey, 1996; Hughes, 1988).

A controversial aspect of washback is whether washback is determined by the test or factors beyond the test itself (Wall et al., 1996). In other word, is it the responsibility of test writers to promote positive washback through guaranteeing test usefulness qualities, or is it interplay of factors which mediate the type and intensity of washback? Empirical studies on washback (Alderson and Hamp-Lyons, 1996; Alderson and Wall, 1993; Hawkey 2006; Qi, 2007; Read and Hays, 2003; Shohamy et al., 1996) indicated that that washback is a complex phenomenon which cannot be considered an automatic or direct effect of exams. Tests' effects may operate quite differently in different situations, and in some situations may not operate at all (Spratt, 2005). Though the connection between testing and learning is commonly made, it is not known whether it really exists and, if it does, it is seldom crystal clear whether tests have favorable or deleterious washback (Hsu, 2009).

These studies also suggested that in addition to test-related factors (e.g. stake, status, purpose, format, content, etc), there are other factors which may be as influential as the test itself, e.g. context-related factors (classroom size, timing of the course, available resource, professional support, etc) and teacher-related factors (e.g. educational background, experience, language ability, training, etc). Thus, washback is not as simplistic as it may seem, rather, it needs to be understood and studied deeply than merely asserted (Alderson and Wall, 1993; Cheng, 2005; Watanabe, 1997).

This study focused on the pivotal role of the teacher as the major agent who can shape positive washback in preparation courses. It deliberately drew on Harlen & Crick's (2003) argument that "Even though there are limits to the action teachers can take to use assessment effectively to help their students learning; they are the only ones whose actions directly affect students" (p.203). Through an ethnographic case study, the teachers' performance were observed and documented at four preparation courses for TOEFL and IELTS.

2. Preparation Courses for High-stakes Testing

High-stake testing which refers to those tests whose results are used to make important decisions about stake-holders (Madaus, 1998) are said to insert powerful effects, either desirable or undesirable, on the classroom context. This study focused on the teaching in the preparation courses for two popular high-stake testes, TOEFL and IELTS.

Ideally, it is expected that preparation courses would mirror a focus on language skills and the practice of appropriate study tasks, such as use of authentic tasks, communicative language teaching and a learner-centered approach which promote learners autonomy (Read and Hayes, 2003, p.5). Though, some researchers believe test preparation courses are to improve test scores by appropriate teaching what is on the test (e.g. Mehrens, 1991), preparation courses are mostly those courses which are set up for training in test taking strategies, familiarizing students with the test, and giving them practice under exam conditions. The available literature on preparation courses mostly show negative assertions about test preparations (Berry and Lewkowitz, 2000; Lumley and Stoneman, 2000). The most controversial aspect of preparation courses is "test preparation practice". Haladyna, Nolen, & Haas (1991, p.4) identified a continuum of preparation from "ethical" to "highly unethical". Ethical practices include activities such as training students in test-awareness skills or attempting to motivate them to perform well. Unethical practices include developing a curriculum which is based on test, preparing teaching objectives which match the test, and practicing past papers. In another effort, Mehrens and Kaminski (1989) suggested preparation continuum consisting of eight forms of preparation, moving from acceptable and ethical practices, such as giving general instruction and test-taking skills, through to the unethical practice of providing instruction on a published parallel form of the test or on the test itself, with a grey area of questionable practices in between.

3. Research Purposes

The major goal of this study was to look closely into the nature of preparation courses to unveil their process as it actually happen. Another catalyst for this study was an underexplored area in literature regarding the role of teachers in preparation courses. The available arguments on the test effects indicate that an exam cannot of itself dictate what and how teachers teach in accountability contexts. Rather, degrees and kinds of test effects occur through the agency of various intervening bodies, with the teacher as the most important one (Spratt, 2005, p.24). There has been no rigorous study to fill this gap by examining how teachers respond to the pressure, mandates and prescriptions of high-stake test contexts and why they do what they do.

Up to now, the available models of washback (Alderson and Wall, 1993; Hughes, 1993, Bailey, 1996) take a top-down flow of effects, from the external test on the teaching process, and little washback research has focused on how teachers, as implementing agencies, could provide bottom-up responses to the tests' effects. This study aimed to develop a tentative model of washback operation in preparation courses.

4. Theoretical Framework

This study drew on concepts from Structuration Theory and Washback Theory. This hybrid approach assisted in examining the process through which high stakes testing influenced teachers' decisions/instruction and teachers' responses to such effects from the tests.

Structuration Theory, which is proposed by Giddens (1984), focuses on the "duality of the institution" via conceptualizing the institution as both constraining and enabling individual actions. It views individuals as knowledgeable agents who interpret, select, and adapt institutional pressures as they take action, but institutional pressures also condition their capacity to interpret and take actions (Scott, 2001). The cognitive approach toward Structuration theory (Coburn, 2004; Spillane, Reiser & Reimer, 2002; Weick, 1995) is based on the assumption that teachers respond to tests' pressures differently because they notice and interpret them differently. It suggests that individuals' knowledge, experience, and context shape their behavior. This conceptualization explains why teachers approach the same test with different pedagogical strategies (Cohen, 1990; Spillane, et al., 2002).

Among the existing washback models, the most appropriate one for this study was BAK model. BAK, which stands for teachers' Beliefs, Assumptions and Knowledge, proposed by Burrows (1998) who examined the washback effect of a classroom-based assessment in the Australian Adult Migrant English Programme. This model further proposes that teachers' responses are related to their attitudes towards and experiences of the implementation of the assessment; their perceptions of the quality of the assessment; the extent to which the assessment represented a departure from their previous practices; and their attitudes to change itself.

5. Research Questions

The first set of research questions dealt with the first objective of the study: how high-stake tests affected the nature of the preparation courses, called hereinafter as process washback:

1. How do different high-stake tests (TOEFL & IELTS) affect content (curriculum/materials) and methodology of the preparation courses?
2. How do different high-stake tests (TOEFL & IELTS) affect the attitudes and perceptions of the participating instructors in the preparation courses?

6. Methodology

The first step in conducting qualitative/ethnographic phase was entering in the site and getting the informants consents (Cohen, et al. 2007, p.171). Individual teachers were given Informed Consent Form which briefly described the overall purpose of the study and data collection. This phase involved an extensive and deep investigation of classroom process. The ethnographer (researcher) engaged in a longitudinal and qualitative observation to record what actually happened in the preparation courses (scope and intensity of tests' washback). The research methodology for this part was ethnographic case study. We used a modified version of an observation scheme, University of Cambridge Observation Scheme (UCOS). The instrument contained lists of text-types used in the classroom and the range of task types according to skill. It also identified teacher initiated, exam-related activities as well as grammar and vocabulary activities. This was followed by the number of times in the lesson when the teachers talked about tests, strategies and the extent to which the teacher adapted the materials to suit the specific requirements of the tests.

6.1. Context

The study was conducted at the Isfahan University Language Center (IULC), a language institute affiliated with the University of Isfahan, Iran. It offered intensive preparation courses in high-stake tests, such as TOEFL and IELTS. The institution was selected because it was accessible and familiar to the investigator and it was also the official representative of English Testing System (ETS), therefore more likely to accept the researcher in their classrooms. The courses were 20-hour courses that ran for two hours in the evening, three days a week, and were organized by skill. Each course was run for specific module. The course could be taken by students of intermediate language level. The maximum class size was 25.

6.2. Participants

This study originally began with a convenient sample of six instructors who were teaching at preparation courses for high-stake tests, TOEFL and IELTS. Although, they agreed at the beginning to be observed, two of them refused to be included in the final report. Due to their own request, the whole data set about them was deleted. Thus, this study only focused on the performance of four instructors: two in IELTS and two in TOEFL PBT preparation courses. All instructors were male and their ages ranged from 30 to 40. Their educational backgrounds were M.A and PhD. In summary, a closer look at the demographic information of four participants (see Table 1) revealed that they were similarly distributed in terms of their ages, academic qualifications, language proficiency backgrounds, teaching General proficiency and high-stake testing.

Table 1. Teachers' Demographic Characteristics

Participants	Age	Degree	Major	Course	General Proficiency teaching experience	High-stake Test teaching experience	Instructor Training Program (for high-stake tests)	Language proficiency certificate
Teacher A	30-35	PhD candidate	ELT	TOEFL	6	3	NO	TOEFL
Teacher B	35-40	M.A. candidate	Literature	TOEFL	10	5	NO	TOEFL
Teacher C	30-35	M.A. candidate	ELT	IELTS	8	4	NO	IELTS
Teacher D	30-35	M.A. candidate	Translation	IELTS	8	5	NO	IELTS TOEFL

7. Results & Discussion

7.1. Ethnographic Observation

UCOS was specially designed to solely focus on the extent to which the course was test-oriented. It had three main areas of focus. First was the analysis of how much class time was spent on activities that were directly related to the test. The types of exam-related texts used in each course were also recorded. There was a separate category which identified any language foci that occurred and categorized them into broad topics. Another reason for using UCOS was examining if the classroom process was ethical or not.

Teacher A:

The total exam-related activity in Class A was around 80%. Teacher mainly talked about proficiency test, went through mock exam papers, and explicitly taught strategy. Talking about proficiency tests was only observed in the first session during which the whole session was devoted to giving complete factual information about TOEFL Reading format, content, timing etc.

The teacher spent a considerable amount of time on analyzing and explaining grammar and vocabulary at the isolated, sentence level (e.g. explicit rules of grammar, synonyms and antonyms, sentence structures, tenses, etc). Only those vocabulary and grammar points on the exam book were covered during class time.

There was only one book (*TOEFL Barrons*) that was used throughout the course, though teacher suggested another commercially published exam book on the first session. The whole classroom procedure seemed to be dictated by the content of the book since the teacher followed every line of the books' instruction religiously.

Teacher B:

In class B, though the teacher focused on vocabulary and grammar, the core of the activities was somehow different from class A, and were not limited to the content of the course book, e.g. activities beyond the isolated sentence-level towards paragraph formation (restatement/paraphrasing, word choice, question structures, underlining words or phrases that contain the main ideas of the text). Furthermore, contrary to class A, his approach toward explanation was implicit and discovery-oriented than explicitly explaining grammar rules. The teacher used three books: a practice- book as the main source ("*Longman Preparation Course for the TOEFL Test*"), a vocabulary ("*504 absolutely essential words*") and a grammar book as auxiliary homework. The teacher did not pay much attention to strategy training.

Teacher C:

In Class C, the teacher spent a greater proportion of time giving information about IELTS and emphasized strategies for effective test-taking. Instructions about test-taking strategies were far more beyond other three courses. In fact, for every activity, the teacher suggested in detail a suitable strategy, and then required the learners to perform similar tasks via employing the same strategy (e.g. efficient use of time, deductive reasoning and guessing, skimming, scanning, and speed reading). Furthermore, most of the time learners were exposed to timed reading mock tests, and were coached on tasks exclusive to the IELTS exam using commercially available IELTS practice materials. The teacher selectively covered three exam-related books during the course ("*Cambridge IELTS*", "*Improve your IELTS Reading skills*") and introduced two exclusive IELTS Strategy Training book ("*101 Helpful Hints for IELTS Academic Module*").

Teacher D:

Teacher D managed his class more skill-based at the early sessions and moved to more exam practice as the course progressed. Negotiation of meaning was more obvious compared to other courses that focused on language analysis as their main activity. He used four sources of textbooks: two IELTS practice book ("*Insight into IELTS*" & "*How to Prepare for IELTS*"), an IELTS strategy-training book, an IELTS exam book which was set as assignment to be completed at home, and a general reading book. He encouraged the students to go through extensive reading on the Internet and sometimes he brought authentic materials from newspapers etc.

Table 2: Exam-related activity and text, UCOS results

	Exam-related Activity	TOEFL		IELTS	
		CourseA	CourseB	CourseC	CourseD
ERA1	talked about proficiency test	First session	Frequently	All sessions	sometimes
ERA21	mock exam papers	Frequently	Two times	Frequently	One session
ERA3	Strategy training	2-3 strategies per session, explicit & decontextualized	Implicit/task-specific	strategy-based training course	Task-specific strategy, integrated
	Exam-related Texts				
ERT1	Commercial Exam textbooks	NO	Yes	Yes	Yes
ERT2	Commercial practice textbook	Yes(Main source)	Yes	Yes	Yes
ERT3	Skills focused textbooks	No	Yes	No	Yes
ERT4	Personally designed authentic materials	No	No	No	Yes

To enhance learners' test wiseness, all teachers referred to IELTS/TOEFL in two ways: Test Information and Test Strategies. They provided the students with factual information about the exam and they gave them information about exam strategies or test-taking tips.

However, the way they covered strategies in the classroom and the amount of the time spent on teaching strategies were quite different in the four courses. Class C was the most strategy-based course and the teacher dynamically embedded every activity in a set of effective strategies, both implicitly and explicitly. On the contrary, teacher B was the most reluctant to go through strategy instruction.

7.2. Washback Evidence from UCOS

This section aims to summarize results from UCOS to examine if the washback of two tests, IELTS and TOEFL, was positive or negative. The evidence of washback was hypothesized as follow:

- Positive: Teacher in preparation courses addressed the future academic needs of the students and helps development of language ability.
- Negative: Teacher delivered the course in such a way just to help the learners to pass the exam and to enhance their score (score-pollution)

These assumptions were made based on Bailey's (1996) criteria for judging whether a language teaching course shows beneficial effects from a test. In her discussion of washback, Bailey (1996, p.260) quotes Morrow (1991) as arguing that "one of the characteristics of a good test is that it should "reflect and encourage good classroom practice" (p,111). For example, if we accept the basic presupposition that IELTS is a modern, communicative test, we might expect it to have the effect of encouraging a communicative, learner-centered type of methodology in both courses C and D. However, this seemed to apply only to course D, but not class C (see Figure1).

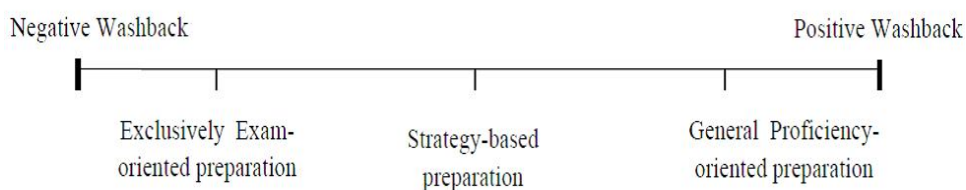
Another clear example could be the amount of lesson time in which the students were actively involved in this type of communication which would be seen as an indication of good classroom practice encouraged by the test for which the students are being prepared. From this view, class D had the highest proportion of communication and negotiation of meaning. Furthermore, the analysis using the UCOS indicated that class A was the most exam-oriented, while class D was the least exam-oriented course, and the teacher embedded the exam-related activities in general proficiency practice.

Another criterion of positive washback was the use of authentic materials (e.g. Messick (1996, p.241) connects authenticity to beneficial washback).

According to Colman (2000, p.35) there is a difference between "test practice materials" and "test preparation materials" which include additional pedagogical guidance for developing language. Results showed that the though teachers were aware of the sources of materials available to them and were free to choose among, their selection was different.

All the teachers made use of the exam-oriented materials, but only Teacher D used materials which was a combination of exam-related, practice-oriented, and some authentic materials he had personally designed for the course. This could be interpreted as an indication of positive washback.

Figure 1. Washback and styles of preparation courses



To examine if these four classes were ethical or not, we follow Mehrens and Kaminski (1989) and Popham (1991) definition of ethical test preparation:

- The test preparation practice should not increase students' test score without simultaneously increasing students' mastery of the content domain being tested

8. Conclusion

This study focused on the pivotal role of the teacher as the major agent who can shape positive washback in preparation courses. Through an ethnographic case study, the teachers' performance were observed and documented at four preparation courses for TOEFL and IELTS. Based on the course description detailed above, it was suggested that the syllabus content and the methodology varied between the exclusively preparation classes (tailoring whole course to meet the TOEFL testing requirements, rather than the students' needs for language development and the learning of academic study skills), to strategy-training class (focusing on familiarizing learners with the test and test-taking strategies, rather than language development) and the classes that received General Proficiency instruction (see Figure 2).

Figure 2. Ethical preparation courses

However, it should be noted that regardless of how “communicative” the design of the test was originally intended to be, its implementation as a required measure of English language proficiency has had the effect of encouraging the development of teaching and learning activities that are narrowly focused on test preparation. Thus, it is unrealistic to expect a preparation course be managed in the same manner as a general proficiency course.

References

- Alderson, J.C. and Hamp-Lyons, L. (1996). TOEFL preparation courses: a study of washback. *Language Testing* 13, 280-97.
- Alderson, J.C. & Wall D. (1993). Does washback exist? *Applied Linguistics*, 14 (2), 115-129.
- Bachman, L.F. and Palmer, A. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Bailey. K.M. (1996). “Working for washback: a review of the washback concept in language testing”. *Language Testing* 13, 3: 257-279.
- Berry, V., and Lewkowitz, J. (2000). ‘Exit-Tests: Is There an Alternative?’ *Hong Kong Journal of Applied Linguistics*, 5(1): 19-49.
- Cohen, L., Manion, I. and Morrison, K. (2007) *Research Methods in Education*. 6th edition. London : RoutledgeFalmer.
- Cheng, L (2004). The washback effect of a public examination change on teachers’ perceptions toward their classroom teaching. In L. Cheng, L. and Y. Watanabe (Eds), *Washback in language testing* (pp.147–170) Mahwah, NJ: Lawrence Erlbaum.
- Cheng, L. (2005). Changing language teaching through language testing: A washback study. *Studies in Language Testing*, 21, Cambridge University Press, Cambridge.
- Denzin, N. K., & Lincoln, Y. S. (2003). The discipline and practice of qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Collecting and interpreting qualitative materials* (2nd ed., pp. 1-45). Thousand Oaks, CA: Sage.
- Green, A. (2007). Washback to learning outcomes: a comparative study of IELTS preparation and university pre-sessional language courses. *Assessment in Education*, 14 (1), 75-97.
- Hughes, A. (1988). Introducing a needs-based test of English language proficiency into an English-medium university in Turkey. In Hughes, A.(Ed.), *Testing English for university study. ELT Documents 127 (pp.134-46)*. Oxford: Modern English Publications
- Hamp-Lyons, L. (1997). ‘Washback, impact and validity: Ethical concerns,’ *Language Testing* 14/3, 295-303.
- Watanabe, Y. (2004). Methodology in washback studies. In L.Cheng, Y. Watanabe & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods*. (pp.19-36). Mahwah, N.J.: Lawrence Erlbaum