Washback of the Public Examination on Teaching and Learning English as a Foreign Language (EFL) at the Higher Secondary Level in Bangladesh

A thesis submitted to the Department of English, Faculty of Arts and Humanities of Jahangirnagar University for the degree of

Doctor of Philosophy
in English (Applied Linguistics and ELT)

Researcher:
Md. Enamul Hoque
Department of English
Jahangirnagar University
Savar, Dhaka, Bangladesh

Supervisor:
Professor Dr. M. Maniruzzaman
Department of English
Jahangirnagar University
Savar, Dhaka, Bangladesh

September 2011

Dedicated to my parents

ABSTRACT

The way in which public examinations influence teaching and learning is commonly described as washback in applied linguistics. Washback influences a variety of teaching and learning areas directly and indirectly, either positively or negatively, or both. The key objectives of the study were to examine washback as a phenomenon relating to those areas that were most likely to be directly affected by the HSC examination in English. The study set out a number of research questions, and answered them to achieve the objectives of the study. The whole study is presented in this thesis divided into six chapters, each chapter incorporating specific issues of the present study.

Chapter One outlines the background information on the general context of the research and identifies the various components of the problem to be studied such as relationship between teaching and testing, statement of the problem, significance of the study, objectives of the study, research questions, definition of terms, limitations of the study, structure of the thesis, etc. Chapter Two covers the theoretical framework of washback relating to the significant areas of the present study. The central issues include the philosophical and empirical bases of testing and washback. Chapter Three presents the literature review of a number of empirical studies carried out on washback in different countries and cultures. The review reveals that washback is a complex phenomenon and has negative or positive relations to teaching and learning EFL. The findings of those studies have shown that, in most of the cases, tests narrow down the syllabus and curriculum, influence the selection of lesson contents, alter teaching methods and materials, but some have indicated that tests have limited or no impact on those areas. Chapter Four discusses the research design and methodology employed in this study. It focuses on how the different types of data were collected, analysed, and presented. A mixed methods (MM) approach was used for data collection and data analysis. The questionnaire (quantitative method), in-depth interview, classroom observation (qualitative method), and analysis of documents were used to collect data. Five-Grade Likert Scale (1932) was used in the questionnaire to elicit responses from the respondents. The subjects, 500 HSC students and 125 English language teachers, were selected

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

i

from 20 higher secondary colleges by using the simple random sampling method. Chapter Five presents the findings, and their interpretation. The Statistical Package for the Social Sciences (SPSS 18.0) was used for quantitative data analysis. Qualitative analysis involved the use of the constant comparative method and inductive logic while quantitative analysis in this study involved descriptive statistics (e.g., frequency counts, means, standard deviations, skewness, kurtosis, median, mode, etc.). The results are sectioned and discussed with reference to the research questions. Chapter Six presents the answers to all the research questions, the findings of the study in brief, recommendations and implications, and the conclusion.

The study indicates that the curriculum corresponds to the textbook, while the EFL public examination does not represent the curriculum and textbook that there is a negative washback of the HSC examination on EFL teaching and learning. The areas mostly influenced by washback were found to be those related to the immediate classroom contexts: (i) teachers' choice of materials, (ii) teaching methods, (iii) classroom tasks and activities, (iv) perceptions of teachers and the learners on the examination, (v) teaching strategies, and (vi) learning outcomes. Based upon the findings, this study put forwards some recommendations for promoting positive washback on EFL teaching and learning at the HSC level. Some of the major recommendations are to: (1) provide testers, examiners, curriculum designers and teachers with extensive professional development opportunities, (2) monitor the teaching and learning activities in the classroom, and check the test related materials whether they enhance EFL learning, (3) align the curriculum and syllabus with the content of the test to assure that students have studied the required contents of the syllabus before taking the tests, and (4) discourage commercially produced clone tests materials.

The study is potentially significant in that it offers educators and policymakers insights into English language teaching and learning at the HSC level. Most importantly, it highlights the voices of teachers and students, the very important people at the centre of the teaching and learning process. It finally advocates the needs for further research on the potential areas of washback.

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

ii

TABLE OF CONTENTS

vii

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

Chapter Three: Literature Review...................................... 107

Chapter Four: Research Methodology.............................. 151

Chapter 5: Presentation and Discussion of the Findings.. 195

xi

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

Chapter 6: Conclusion

References.... 409

LIST OF TABLES

xiv

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

LIST OF FIGURES

xvi

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

ABBREVIATIONS AND ACRONYMS

AILA	-	Association internationale de linguistique appliquée *or* International Association of Applied Linguistics
ANOVA	-	Analysis of Variance
ASL	-	Arabic as a second language
ATESL	-	Administrators and Teachers of English as a Second Language
BAK	-	Beliefs, Assumptions, Knowledge
BAKE	-	Beliefs, Assumptions, Knowledge, Experience
BISE	-	Board of Intermediate and Secondary Education
CEELT	-	Cambridge Examination in English for Language Teachers. Tests the English competency of non-native teachers of English
CET	-	College English Test
CEIBT	-	Certificate in English for International Business and Trade for advanced levels.
CPE	-	Certificate of Proficiency in English
CLA	-	Communicative Language Ability
CLT	-	Communicative Language Teaching, a teaching approach of second and foreign languages that emphasizes communication and interaction as both the means and the goal of learning a language.
COLT	-	Communicative Orientation to Language Teaching
DA	-	Discourse Analysis
DELNA	-	Diagnostic English Language Needs Assessment
DSHE	-	Directorate of Secondary and Higher Education
EAL	-	English as an Additional Language
EAP	-	English for Academic Purposes
EFL	-	English as a Foreign Language
ECCE	-	Examination for the Certificate of Competency in English (Michigan University) - lower level.
ECPE	-	Exam for the Certificate of Proficiency in English (Michigan University) - higher level.
EFT	-	English For Today
EGP	-	English for general purposes
ELD	-	English Language Development

xix

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.　　<Pages 288-786>

ELP	-	English Language Portfolio
ELT	-	English language training or teaching
EIP	-	English as an International Language
ELTIP	-	English language Teaching Improvement Project
EPTB	-	English Proficiency Test Battery
ELTS	-	English Language Testing Service
ESL	-	English as a Second Language
ESAP	-	English for Specific Academic Purposes
ESP	-	English for Specific Purposes
ESOL	-	English for Speakers of Other Languages
ETS	-	Educational Testing Service
FCE	-	First Certificate of English
FFPS	-	Full-fee Paying Students
FL	-	Foreign Language
FLA	-	Foreign Language Acquisition
GPA	-	Grade Point Average
GMAT	-	Graduate Management Admission Test.
GPA	-	Grade Point Average
HSC	-	Higher Secondary Certificate
IELTS	-	International English Language Testing System
IATEFL	-	International Association of Teachers of English as a Foreign Language
IDP	-	International Development Program
L1	-	Language 1 - native language
L2	-	Language 2 - the language we are learning
LL	-	Language Learning
LSP	-	Language for Specific Purposes
M	-	Mean Score
MCQ	-	Multiple Choice Question
MDI	-	Measurement Driven Instruction
MANOVA	-	Multivariate Analysis of Variance
MoE	-	Ministry of Education
MSE	-	Mean Squared Error
MT	-	Mother Tongue
MTELP	-	Michigan Test of English Language Proficiency

xx

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.　　<Pages 288-786>

NCTB	-	National Curriculum and Texbook Board
NNL	-	Non-Native Language
NNS	-	Non Native Speaker
NS	-	Native Speaker
OET	-	Occupational English Test
S	-	Student
SL	-	Second Language
SLA	-	Second Language Acquisition
SSC	-	Secondary School Certificate
STDV	-	Standard Deviation
T	-	Teacher
TEFL	-	Teaching English as a Foreign Language
TEIL	-	Teaching English as an International Language
TESOL	-	Teaching English to Speakers of Other Langauges
TOEFL	-	Test of English as a Foreign Language
TOEIC	-	Test of English for International Communication
TESL	-	Teaching English as a Second Language
TLU	-	Tasks in the Language Use
UCLES	-	University of Cambridge Local Examinations Syndicate
UCOS	-	The University of Cambridge Classroom Observation Schedule
UEE.	-	University Entrance Examination
VE	-	Vocational English

Chapter One
Introduction

The first chapter offers an introduction to the context of the whole study by giving a brief account of the underlying problems that generated this research study. This chapter consists of nine sections and provides an overall introduction to the study. It incorporates a number of issues, and identifies various components of the problem to be studied including the background information on the general context of the research, relationship between testing, teaching and learning, testing at the higher secondary level in Bangladesh, importance of studying washback, statement of the problem, objectives of the study, significance of the study, research questions, definition of terms, limitations of the study, structure of the thesis, and a conclusion.

1.1 The General Context of the Research

The Bangladesh education system is characterised as being examination-driven. Under this system, examinations are of exaggerated importance. At various levels of education, be they secondary, higher secondary or tertiary, it is a common practice that teachers teach to the test. Not only are most courses tailored to examinations, but the teachers' and students' attention is also correspondingly directed at the skills which will be tested in the examination. Furthermore, test scores are viewed both as a marker of students' academic success and as the premise to their future career. Testing is generally accepted as an integral part of teaching and learning. It is one of the basic components of any curriculum, and plays a pivotal role in determining what learners learn. Tests also play a central role in deciding on what to teach, and how to teach. Candlin and Edelhoff (1982) assert that learners learn most when they are quite precisely aware of how their efforts are to be judged and evaluated.

It has long been widely recognised that a high-stakes test such as the HSC public examination can have a major impact on educational systems and on the societies. Pearson (1988) points out that "public examinations influence the attitudes, behaviours, and motivation of teachers, learners and parents, and because

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

1

examinations often come at the end of a course, this influence is seen working in a backward direction, hence the term 'washback'" (p. 98). In addition, washback has been generally perceived as being bipolar – either negative (harmful) or positive (beneficial). The research investigated washback of the HSC examination on teaching and learning English as a foreign language.

In this study, the terms *assessment* and *test* are used interchangeably with *examination*, as has also been done in educational literature. Although assessment has also come to include the evaluation of schools or education systems, this aspect will not form a part of the following discussion. Here, the primary focus of the discussion will be on what is commonly termed, "high-stakes examinations". Assessment is often called "high-stakes" if it has real or perceived effects on the life or academic opportunities of students and consequences for teachers and schools. The term "public examination" is synonymous with an external examination or a test that is administered by external agencies or forces to evaluate learning products or results with a decisive consequence or influence on test-takers (Alderson, 1986; Shohamy, 1992).

Generally, public examinations are held at the state level at the end of academic years, and controlled and administered by external examining boards. These are academic achievement tests. Examination boards are designated for conducting the examinations and issuing certificates through assessment of answer scripts. Education boards are formed with the main objectives of maintaining standards of education. The Higher Secondary Certificate (HSC) examination in Bangladesh, the subject of the present study, is a high-stakes test. It is an external test because it is administered by an external body, an education board. It can also be termed as a standardized test because of its nature; it is a criterion-reference test in characteristics as well.

Traditionally, the HSC examination can be termed as an achievement test. The relationship between testing and teaching has long been a matter of interest in both educational and applied linguistics. In applied linguistics, the influence of testing on teaching and learning has been referred to as *washback*. A "high-stakes" test can directly and powerfully influence how teachers teach and students learn. Testing is often seen as both a necessary evil and a vehicle for effecting educational change, especially when the educational system is driven by tests or examinations.

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

2

High-stakes tests influence the contents and methodology of teaching programmes, attitudes towards the value of certain educational objectives and activities, the academic employment options that are open to individuals, and may have significant long-term implications for education systems and the societies in which they are used. In Bangladesh, English language is taught compulsorily as a study subject in the higher secondary education. It is taught as a foreign language (EFL), and practiced within a context-restricted environment in which the determiners of language learning phenomenon depend on classroom activities, determined by the classroom teacher. HSC level students study English subject comprising two papers carrying 200 marks; and they sit for the public examination at the end of two years of study. It is often assumed that washback exists to influence teaching and learning to a certain extent. So, it needs to examine whether this public examination influences English language teaching and learning. This influence is termed as washback which may be positive or negative towards language teaching and learning. Hence, it is very crucial to find out which aspect of washback dominates English as a foreign language (EFL) teaching and learning at the HSC level in Bangladesh.

1.1.1 Teaching and Testing EFL at the Higher Secondary Level

The HSC English is based on the communicative approach to teaching a foreign language, and emphasises students' communicative competence. The course is supposed to prepare students for real-life situations in which they may be required to use English. The selection of the course content has been determined in the light of students' present and future academic, social, and professional needs. The HSC examination is an achievement test. Although it refers to the syllabus, it seldom takes teaching contents into consideration. This causes the separation of tests from teaching the syllabus, which, in turn, causes students to value tests more than regular class performance. Many students think that so long as they can pass the test it does not matter whether they attend the regular classes or not; this results in students' high rate of absence from classes in some colleges, especially in the rural areas.

The major part of the present HSC examination is mainly composed of vocabulary items, matching, rearranging, grammatical, cloze test questions, and restricted composition items. It has been proved by evidence that students who take

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

3

these types of tests can significantly increase their scores "artificially" (Alderson et al., 2001, p.45). This encourages both teachers and students to work over test skills and countermeasures in preparing for the test, which interferes in regular classroom teaching, leads to test-oriented teaching, and consequently affects students' systematic mastery of the fundamental knowledge and integrated skills of English, and hinders students' development of communicative competence.

Those who set question papers may be academically highly qualified, but hardly have any training in question paper setting and modern approaches to assessment. The examiners do not receive any formal guidelines for scoring/evaluation of the answer scripts; they prefer to check scripts as quickly as possible. The question papers are hardly representative of the entire curriculum. Teachers and students mostly rely on guidebooks, model questions, and suggestions book for the preparation for the examination. The prescribed textbooks are hardly followed. Examination questions are repeated in at least every two or three years, and hence questions can be predicted. There are 'model question papers', or 'guide books' available in the market with ready-made answers based on recently past years' questions. Teachers and students tend to rely on such guides and put their content to memory. The HSC examination, thus, has become a dreadful thing and an end in itself rather than a means to achieve educational objectives of improving teaching and learning and raising standards and quality of education. Students are fearful of examination, and at times unsuccessful students commit suicide.

In the twenty-first century, many countries are increasingly confronted with rapid social, economic and political changes that take place in their societies as a result of technological innovations and the process of globalisation. These nations often turn to their educational system to help prepare their youth and citizens for the challenges that they must face. As a result, the authorities are becoming increasingly aware of the needs to reform educational practices to bring them in line with the realities and demands of a new age. In Bangladesh, education is regarded as a vital tool in the task of social advancement, preparation of human resources and social engineering.

The education system in Bangladesh is presently undergoing a reform that includes syllabuses and curriculums, examinations, textbook materials, organisational and responsibility changes. The HSC English curriculum and syllabus

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

4

developed in 1990 had been under serious criticism for not providing an adequate level of basic oral-aural communication competences for the higher secondary students though they had studied English for twelve years. The government, therefore, undertook initiatives for the revision of the old HSC English syllabus and curriculum, and initiated for the writing of new textbooks with communicative view of teaching and learning.

The new curriculum for the HSC EFL education was introduced in 2000, following by the issuance of the new textbooks to be used by the students from 2001. Under the present syllabus and curriculum, the first HSC examination in EFL was held in 2003. English *second paper* was modified in 2007 (examination was first held nationally in 2009) introducing more grammar, composition (subjective question) and some textual items based on the new requirements. However, the idea that the reform can encourage student learning in a qualitative way has yet to be attested empirically. The foreign language test was designed to replace the old elective test that was mainly oriented towards the evaluation of grammatical aspects of the foreign language education. The new test seeks to evaluate the communicative competence of the students, which means observing those aspects in which they can use the knowledge they possesses of that foreign language to act in specific situations which demand their making use of that knowledge. The new English curriculum developed by the NCTB as a framework for the examination makes explicit opportunities of the Communicative Language Teaching (CLT) approach as the official orientation of the teaching of languages in the country based on Littlewood (1981), Widdowson (1978), Brumfit and Johnson (1979), Halliday (1970), Hymes (1972), Canale and Swain (1980), and Canale (1983) among others.

It is widely believed that when designing a language test or evaluating its potential usefulness, two critical measurement qualities are needed to give consideration to: reliability and validity. Validity relates to the extent to which meaningful inferences can be drawn from test scores (Bachman, 1990). In contrast, reliability concerns the consistency of measurement. Of the validity considerations for a language test, construct validity is viewed as pivotal. It is often used to refer to the extent to which one can interpret a given test score as an indicator of a test takers' language ability. The term can be interpreted to mean that if a test has good construct validity, it is a good indicator of test takers' language ability and vice-

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

5

versa. Bachman and Palmer (1996) place special emphasis on test tasks claiming that they should be carefully selected and their characteristics should be adequately described. Construct definition is given by Chapelle (1998) as a theoretical description of the capacity that a test is supposed to measure. Bachman and Palmer (ibid.) seem to suggest that the more the test tasks reflect the construct definition, the higher the construct validity. From their perspective, construct validity is affected to some extent by the characteristics and content of the test tasks. In this regard, there is an obvious need to examine the task characteristics of the HSC examination in English. Despite the present syllabus being communicative, it is observed that, there is not sufficient evidence of teaching the two important skills: listening, and speaking. Testing of listening and speaking are ignored in the examination. Therefore, teachers of English consider teaching listening and speaking simply waste of time. The contents of the HSC examination in English can hardly assess students' communicative competence. Therefore, the validity of the HSC examination in English is doubtful in term of testing communicative competence.

Eight general secondary and higher secondary education boards are designated to administer the examinations and issue certificates. Different education boards conduct the examination with separate sets of question papers under the same syllabus and textbook. The question format, pattern, contents of test, and the distribution of marks for the tasks and items are same in nature in all boards. In the field of communicative language testing research and practice, the framework proposed by Bachman and Palmer (1996) is often taken as a theoretically grounded guideline for analysing the characteristics of a test.

The HSC examination in English does not correspond to the curriculum objectives. They contain little reference to the knowledge and skills that students need in their everyday life outside the class, and they tend to measure achievement at a low taxonomic level. As can be seen from the discussion above on the purposes for which examinations are used in educational systems and the support or critique surrounding it, this is an issue that is still widely debated. Be that as it may, as Cheng (2004) points out, teaching and testing will probably become more closely linked in more complex manners in the future. In Hong Kong, Andrews et al. (2002) have found that during the last four years of secondary schooling, the focus is still on preparing students to pass the mandatory public examinations; in fact, the

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

6

developmental work on the new English Language syllabi in Hong Kong deliberately targets a positive washback effect of the examination on classroom teaching. Therefore, it is important to conduct studies that examine what is actually happening in schools and classrooms, because, as mentioned in Wall (1997) and Bailey (1996), the claims about test consequences are sometimes based more on assumptions than on empirical evidence. As such, studies which provide empirical evidence showing how innovations in exams or testing affect teaching and learning in classroom are crucial to validating these claims.

While HSC examinations have promoted the college English teaching, they have also led to the test-oriented teaching in colleges and hindered the development of students' communicative competence. To eliminate the negative washback effect of the HSC examination in English, subjective questions are increased. However, it is found that teachers still mainly adopt traditional methods to teach writing and tend to ignore the intention of reform. Teachers' beliefs and experiences in language teaching are found to be one of the contributing factors. Another factor is that the status of teachers in Bangladesh is very much related to the test scores achieved by their students. Teachers' perceptions of an examination can be as significant as the test itself. It was a pressing need to examine how the HSC examination in English influenced the academic behaviours of both the teachers and the students. Therefore, the present study attempted to investigate the influence of the HSC examination on EFL teaching and learning.

1.1.2 Importance of Studying Washback

The strong influence of high-stakes tests on teaching and learning processes has long been accepted in the field of education. In the field of applied linguistics, the concept of a test influencing teaching and learning in the language learning classroom was rarely discussed until the early 1990s (Andrews, 2004; Bailey, 1996; Cheng, 1997; Elder & Wigglesworth, 1996; Wall, 2000). The term *washback* became used in the field to refer to the power that high-stakes tests could have on language teaching and learning, although *impact* or *consequences* are more commonly used in the field of education (Rea-Dickins & Scott, 2007). While the concept of washback was earlier only asserted based on anecdotal evidence (Burrows, 2004), the pioneer evidence-based washback research was carried out by

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

7

Alderson and Wall (1993). They investigated the effects of the introduction of new tests in Sri Lanka on the teaching of English as a foreign language by secondary school teachers. Implementing the tests was expected to reinforce innovations in teaching materials and to encourage communicative language teaching while discouraging traditional grammar focused teaching. They found, however, that teachers' lessons remained teacher-centred over the period of two years and students still had little chance to use English in a practical way – although language learning activities and the design of classroom tests were influenced by the new textbooks. They concluded that the effects of the implementation of new tests were much more limited than expected and that the mechanism of washback was not as straightforward as previously thought.

When studying washback, it is also possible to focus on participants (teachers, students, material developers, publishers), process (actions by participants towards learning), and products (what is learned and the quality of learning), as suggested in Hughes's trichotomy model (Hughes, 1993 as cited in Bailey, 1996). Watanabe (2004) proposes disentangling the complexity of washback by conceptualizing it in terms of: Dimension (specificity, intensity, length, intentionality and value of the washback), aspects of learning and teaching that may be influenced by the examination, and the factors mediating the process of washback being generated (test factors, prestige factors, personal factors, macro-context-factors). Usually researchers focus on one aspect or type of washback. In Alderson and Wall's study in Sri Lanka (Alderson & Wall, 1993; Wall, 1996), the introduction of a test of English as a foreign language proved to produce faster changes in the content of teaching than changes in teaching methodology. Cheng (1997), in the preliminary results of a study of the washback effect of the Hong Kong Certificate of Education Examination in English in Hong Kong secondary schools, reports that washback effect "works quickly and efficiently in bringing about changes in teaching materials […] and slowly and reluctantly and with difficulties in the methodology teachers employ" (p.1). Cheng introduces the term 'washback intensity' to refer to the "degree of washback effect in an area or a number of areas that an examination affects most" (p.7).

Andrews et al. (2002) finds out in their study that the impact of a test can be immediate or delayed. According to these researchers, washback seems to be

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

8

associated primarily with 'high–stakes' tests, that is, tests used for making important decisions that affect different sectors., for example, determining who receives admission into further education or employment opportunities (Chapman and Snyder, 2000). Madaus (1990 in Shohami, Donitza-Schmitdt & Ferman, 1996) identifies as 'high' such situations when admission, promotion, placement and graduation are dependent on the test. Cheng (2000) reports on how tests are often introduced into the education system to improve teaching and learning, especially in centralised countries including China, Taiwan, Japan, and Hong Kong where tests are considered an efficient tool for introducing changes into an educational system without having to change other educational components. In some countries these tests can be considered "the engine for implementing educational policy" (Cheng, 2000, p. 6).

In recent years, researchers have been making significant inroads into investigating this phenomenon in different social and educational contexts. As a result, the definition as well as the nature and scope of washback have been extensively discussed, and a number of different perspectives have emerged in language testing and ELT research area. Despite the strong link between testing, teaching and learning discussed in the field of education, the assertion that a test influences what teachers and students do in the classroom is often based on anecdotal evidence, and did not receive much attention from researchers until the early 1990s in the field of applied linguistics.

Of the various patterns or themes that have emerged from studies on washback, the most prominent one is the gap that exists between teachers' beliefs about innovation and the beliefs held by innovators. There is sufficient evidence indicating that teachers' perceptions of washback seldom overlap the perceptions of test designers or policy. Though some research studies have been carried out on washback studies in different countries, no formal research has been carried out in Bangladesh to investigate how the HSC examination has been influencing EFL teachers and students. So, it was important to carry out a study in Bangladesh on this topic. Furthermore, though a good number of washback studies have been carried out during recent years in different countries, the washback effect is still to be adequately defined and analysed.

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

9

1.1.3 Relations of Testing to Teaching and Learning

There is an in-depth relation between testing, teaching and learning. Test objectives determine the teaching objectives. Testing strongly influences the classroom activities. Tests are assumed to be powerful determiners of what happens in classrooms; and it is commonly claimed that tests affect teaching and learning activities both directly and indirectly. As mentioned earlier, washback, a term commonly used in applied linguistics, refers to the influence of language testing on teaching and learning. The influence of a test on the classroom is, of course, very important; washback effect can be either beneficial or harmful. Teachers as well as their students tailor their classroom activities to the demands of the test, especially when the test is very important for the future of the students.

A high-stakes test is a type of test whose results are seen- rightly or wrongly by students, teachers, administrators, parents, or the general public as the basis upon which important decisions that immediately and directly affect the students are made. A test can be considered as high-stakes if the test results are perceived by stakeholders (e.g., teachers, students, parents and schools) to have serious consequences, such as graduation, comparison or placement of students, the evaluation of teachers or schools, and/or the allocation of resources to schools (Madaus, 1988). High–stakes tests can be norm or criterion–referenced, and internal and external in origin. They offer future academic and employment opportunities based upon the results. They are usually public examinations or large-scale standardized tests. The HSC public examination, the subject of the study, is such a high-stake test. It is given to the students at the end of their 12th year of education. Students either proceed to further studies or leave school, and seek employment after passing the HSC examination.

Washback is the power of examinations over what takes place in the classroom (Alderson and Wall, 1993, p.115). Numerous explanations of the term 'washback' can be found throughout the published research and literature on language testing. One of the most common definitions sees the concept referred to as the influence of testing on teaching and learning (e.g. Alderson & Wall, 1993; Gates 1995; Cheng & Curtis 2004). Brown (2000) defines washback as "the connection between testing and learning" (p.298). Gates (1995) explains washback simply as "the influence of testing on teaching and learning" (p.101). Messick (1996) refers to

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

10

washback as "… the extent to which the introduction and use of a test influences language teachers and learners to do things they would not otherwise do that promote or inhibit language learning" (p. 241).

Pierce (1992) states that the washback effect is sometimes referred to as the systemic validity of a test. Bachman and Palmer (1996, p.29-35) have discussed washback as a subset of a test's impact on society, educational systems, and individuals. Alderson and Wall (1993) consider washback as the way that tests are perceived to influence classroom practices, and syllabus and curriculum planning. Cohen (1994) describes washback in terms of how assessment instruments influence educational practices and beliefs. Public examinations are often used as instruments to select students as well as a means to control a school system, and are commonly believed to have an impact on teaching and learning. Given that external tests or public examinations have exerted an influence on teachers and students with an associated impact on what happens in classrooms, such a phenomenon is denoted as "washback" or "backwash" (Alderson, 1986; Morrow, 1986; Pearson, 1988; Hughes, 1989; Morris, 1990). As tests have the power to select, motivate and reward, so too can they de-motivate and punish.

Language tests have become a pervasive part of education system and civilization. They play a significant socio-economic role in modern societies. A test is an experience that the teacher creates to serve as a basis for grading a learner in order to group them according to a laid down standard by a government or an institution. A test is a method that generally requires some performance or activity on the part of either the testee or the tester, or both. There is a set of techniques, procedures, test items that constitutes an instrument of some sort. Such a type of external test is commonly believed to have an impact on teaching and learning. Every test does not carry the same weight and importance. High stakes tests influence the way students and teachers behave, the content and methodology of teaching programmes, attitudes towards the value of certain educational objectives and activities, the academic employment options that are open to individuals, and may have significant long-term implications for education systems and for the societies in which they are used.

According to Alderson and Wall (1993), the notion that testing influences teaching is referred to as 'backwash' in general education circles, but it has come to

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

11

be known as 'washback' in applied linguistics (p. 11). Washback and backwash are now interchangeably used in both EFL and ESL research of applied linguistics (Bailey, 1999)**.** Washback or backwash has been defined as "a part of the impact a test may have on learners and teachers, on educational systems in general, and on society at large" (Hughes, 2003, p. 53).

In recent years, there has been growing interest among the testers in the field of education, in the effects, both desirable and undesirable, of tests and the concepts of 'test impact' and 'test washback'. Impact is the consequence of a test on individuals, on educational systems and on society in general. The term 'washback' or 'backwash' as it is sometimes refers to, can be broadly defined as the effect of testing on teaching and learning, and is therefore a form of impact. It is then a concept which includes several specialised areas in the field of applied linguistics such as communicative language teaching and testing.

The term "Communicative language teaching and testing" has emerged as a much-talked issue in the worldwide English language education arena. In communicative language teaching, the purpose of testing is to evaluate how far learning and teaching are taking place, or in other words, how far the students have attained the ability to use the language in certain span of time. Testing communicative competence means testing the ability to use language for communication. This also includes the testing of four basic skills of language — listening, speaking, reading, and writing. There are some distinctions between the traditional examination and the communicative language test. The purpose, in traditional examination, is that of promoting or detaining a student, or awarding degree. In language tests, how far the learners have attained language proficiency has to be measured.

There is a natural tendency for both teachers and learners to tailor their classroom activities to the demands of the test, especially when the test is very important for the future of the learners, and the pass rates are used as a measure of teachers' success. There is a consensus among the educators that the contents of classroom instruction should be decided on the basis of clearly understood educational goals, and examinations should try to ascertain whether these goals have been achieved. The influence of examinations on second/foreign language (SL/FL) teaching and learning has become an area of significant interest for testers and

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

12

teachers alike. Negative washback is said to create a narrowing of the curriculum in the classroom so that teachers and learners focus solely on the areas to be tested. On the other hand, there have been attempts to generate positive washback by means of examination reform to encourage teachers and learners to adopt more modern communicative approaches to language learning. When the examination does that, it forces learners and teachers to concentrate on these goals; and the washback effect on the classroom is very beneficial.

Testing has been used for decades, but concerns about its influence have recently increased. Davies et al. (2000) define 'impact' by as "the effect of a test on individuals, on educational systems and on society in general" (P. 79). With this increased concern, the influence of tests has been officially termed as 'washback' or 'backwash', and used as an impact in the field of language testing. Washback appears to be a concern in education in general. This study, however, focuses on washback on SL/FL education. Specifically, the EFL test in the Higher Secondary Certificate (HSC) examination is the subject matter of the present study.

1.2 Statement of the Problem

Examinations play an important social and educational role in Bangladesh; the promotion of an effective English testing system has thus been of great importance. It is now widely believed that the phenomenon of how external tests/public examinations influence teaching and learning is commonly described as "washback" in language instruction. Literature indicates that testing washback is a complex concept that becomes even more complex under a variety of interpretations of the washback phenomenon on teaching and learning. Some studies conclude that no simple washback effect occurs (Alderson and Hamp-Lyons, 1996; Watanabe, 1996), whereas others find powerful determiners of language testing towards classroom teaching (Hughes, 1988; Khaniya, 1990; Herman and Golan, 1991).

Testing and teaching are strongly correlated; testing determines the teaching and the learning. Testing objective is determined by teaching objectives. Teaching and learning in Bangladesh are test-driven. The classroom activities are overwhelmingly guided by the contents of the examination. The teachers remain very selective to the classroom activities. They teach those items which are likely to be tested, and ignore the ones that may not be tested. They narrow down the syllabus

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

13

for the benefit of the test. They directly teach to the test to attain the immediate goal of scoring high in the examination. There exist mismatches between the curriculum objectives and examination objectives. There is strong disagreement on whether all the skills of English language are properly tested in the public examinations. These are all problems to be addressed in the present study.

Most recently, policy makers, educators and researchers in Bangladesh have devoted much effort to the nature and outcomes of the examination and its washback on EFL teaching and learning. Though considerable amounts of washback studies have been conducted in various contexts of English language teaching and learning through out the world during the last decade, a little research has been carried out within Bangladesh context. Therefore, the present study was designed to examine the washback of the Higher Secondary Certificate (HSC) examination on EFL teaching and learning.

The present study may be taken as a pioneer formal research on the washback of the public examination on EFL teaching and learning at all levels in general, and the HSC in particular. The study investigated the relationships between the EFL curriculum and examination, the textbook materials and EFL test, teaching method and EFL examination, classroom activities and test, etc. Then the study examined whether any washback of the HSC examination existed, and how much and in what way teaching and learning English were influenced by the HSC examination.

1.3 Objectives of the Study

The present study entitled *Washback of the Public Examination on Teaching and Learning English as a Foreign Language (EFL) at the Higher Secondary Level in Bangladesh* was designed to examine whether the washback of the HSC public examination influenced teaching and learning English as a foreign language as a whole. The study investigated the phenomenon of the washback effect in the light of measurement driven instruction. The present study tried to understand how the main participants in the Bangladesh educational context react to the HSC examination– a major public examination. It attempted to explore the nature and scope of the washback effect on the aspects of institutional policies, teachers' and students'

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

14

perceptions, and teachers' behaviours, within the context of the HSC examination in English.

Testing and standards appear to be a permanent part of today's educational arena; and, since teachers are obliged to work under these guidelines, it is important for educational research to examine if testing and standards influence teachers' activities in the classroom. Teachers are at the center of this debate, and have a vested interest in its outcomes. Therefore, the purpose of this study was to determine the existence and the degree of this influence. The study hoped to learn whether testing changed the teachers' teaching methods; whether the teachers were influenced to change their beliefs, strategies and activities to align the test.

The present study explored the possible answers to all the research questions posed in the present study. This study attempted to find out whether the EFL teachers are truly teaching to the test and the potential reasons involved. The broad purpose of this study was to investigate how those were involved, directly and indirectly, in teaching and learning English. The purpose of this study was also to determine in what ways the teachers followed the syllabus and curriculum and teaching method to impact test results, to what degree, and in what specific way it was done. On the whole, all the conclusions were drawn based on what the teachers, the students, and the other participants said in the present study. Thus, the objectives of the present study can be summarised as follow:

General objective

The study was designed to generally investigate how the HSC examination in English directly and indirectly influenced teaching and learning English as a foreign language (EFL).

Specific objectives

The study specifically:

a. explored the nature and scope of the washback effect on the aspects of teachers' and students' perceptions and behaviours within the context of the HSC examination in English;

b. tried to understand how the main participants (e.g. students, teachers, examiners, curriculum specialists and materials writers, and the like) within the Bangladesh education context react to the examination in EFL;

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

15

c. intended to determine the ways teachers and students follow the syllabus and curriculum, textbook, materials, etc;

d. attempted to investigate whether the objectives of the syllabus and curriculum are achieved through classroom teaching and learning; and

e. endeavoured to learn whether the examination influence changes teachers' teaching methods, teaching strategies and activities to align the test or curriculum.

1.4 Significance of the Study

The strong influence of high-stakes tests, such as the HSC examination, on teaching and learning process has long been accepted in the field of education. The study proved to be highly significant in many respects: the study examined the influence of public examination on teaching and learning English as a foreign language; the findings of this study would provide educational parties involved in English language education with important information to help improve the policy, practice and implementation of English language teaching and learning. Most importantly, the study highlighted the voices of teachers and students, the very people at the centre of the teaching and learning process.

Despite testing being a very important activity in the teaching learning situation, a little formal research (e. g. Maniruzzaman and Hoque, 2010; Maniruzzaman, 2011) has been carried out in the field of washback effect till today in Bangladesh. Though some research studies have been carried out on washback studies in different countries, to my knowledge, no research has been carried out in Bangladesh to investigate how the HSC examination has been influencing EFL teachers and students. So, there were ample scope of study in this filed. It was expected that the study would bring a quality change in the present examination system at the higher secondary level in Bangladesh. The study may help teachers and students consider the examination as 'servant' to the learning, not the 'master', 'lever', not the 'barrier'.

One of the main strengths of this study is its research design. The study used a mixed- methods (MM) approach to both data collection and analysis. The results were relatively greater in breadth and depth, not only in terms of the data collection but also in terms of the interpretation. Therefore, the results may be considered

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

16

reliable. This was an empirical study; and it was one of the few washback studies that employed both quantitative and qualitative data to explore the washback effect on teaching and learning. Based on both quantitative and qualitative data, this study provided solid research evidence to describe and explain the washback effect of the HSC examination in English subject on various aspects of teaching and learning, and on the Bangladesh education system as a whole. Although this investigation provided data on and evidence of the washback effect in a specific educational context, it should also contribute to the understanding of education in Bangladesh, in general.

The questions asked in the various instruments- questionnaires, interviews and the classroom observation scheme, have drawn on theoretical considerations in the areas of language teaching and learning along with interviews with relevant stakeholders in Bangladesh. The instruments are, therefore, easily applicable to future studies conducted at other levels of education in Bangladesh. It is believed that the study provides a starting point for future researchers to find the most appropriate method for their own contexts. It would facilitate further research on washback, and allow easier comparison of the results between the studies.

The study was potentially significant as it offers educators and policymakers insights into English language teaching and learning at the HSC level. The study, first, investigated the relationships among the curriculum, the textbooks and materials, the EFL teaching and learning, and the HSC examination in English; the study then tried to explore whether the HSC examination exerted any washback on the EFL teaching and learning. This study further discerned the nature of washback and the variable(s) influenced by the washback effect. The findings of this study may provide important information to help the educational parties involved in English language education modify the policy, practice and implementation of any innovations for the improvement of English language teaching and learning.

The results of the study may enormously contribute to the area of marking and grading of the students in the examination. The findings of the study may contribute to the literature on general education, EFL teacher education, and cognitive psychology.

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

17

1.5 Research Questions

Washback is a very complex notion. It does not only refer to the effect of an examination in the classroom, but also in the school, in the educational system and in the society. It is simplistic to believe that a test can result in all desired changes in teaching and learning. In the public examination system in Bangladesh, some language elements (vocabulary, grammar, etc.) and the two literacy skills: writing and reading comprehension are tested in the examinations, while the other two language skills: listening and speaking remain entirely untested. It is now a proven belief that activities in schools are dictated by examinations (Wong et. al. 2000). When examinations are high-stakes tests, their impact is maximised. Moreover, changes in education, particularly in teaching can be facilitated by tests (Davies, 1985).

Based on the research purposes, the study looked at the washback effect of the HSC examination on EFL teaching and learning both at the macro level with respect to major parties within the Bangladesh educational context and at the micro level with regard to different facets of classroom teaching and learning. It is important to emphasise that both teaching and learning were studied in this project, as both of these constructs occur interactively in the classroom. Therefore, the teachers and the students were included in the study. However, aspects of learning and learners were studied only when they related to classroom teaching. Washback researchers in the field of applied linguistics have rarely communicated with those in the field of education, although the power that a test has on teaching and learning is now well recognised and has been extensively investigated in both fields (Rea-Dickins, 2004). Therefore, the current research aimed to incorporate theories of test impact or washback available in the two fields (education and applied linguistics). It can be argued that identifying the role of contexts and beliefs can contribute to a model of washback which shows further understanding of its mechanism.

The study focuses on observing what happens in language classes for preparing the students for the HSC examination. Due to the scarcity of similar studies of washback, the objectives of the study were methodological as well as substantive. Students tend to be influenced by their teachers in terms of the relationships between teaching and learning; nevertheless, students' views may be different from, or independent of, their teachers'. For this reason, the present

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

18

researcher focused on both teacher and student perception, and compared both of them in order to look at how differently they think and feel about the influence of the HSC examination on teaching and learning. Therefore, the study posed the following research questions:

RQ$_1$. Does washback of the HSC examination influence EFL teaching and learning?

RQ$_2$. Does the HSC examination have any washback effect on the syllabus and curriculum?

RQ$_3$. To what extent does the test content influence teaching methodology?

RQ$_4$. What are the nature and scope of testing the EFL skills of the students at the higher secondary level?

RQ$_5$. What are the effects that an examination preparation process can have on what teachers and learners actually do?

RQ$_6$. What is the effect of the HSC examination on the academic behaviour, feelings, perception and attitudes of teachers and students?

1.6 Definition of Terms

A number of key terms are defined as follows in order to establish a consistent and common meaning for them as they are used in this thesis.

Achievement Test: An achievement test measures what a learner knows from what he/she has been taught. This type of test is typically given by the teacher at a particular time throughout the course covering a certain amount of material.

Alternative Assessment: Alternative assessment refers to a non-conventional way of evaluating what students know and can do with the language. It is informal and usually administered in the class. Examples of this type of assessment include self-assessment and portfolio assessment.

Analytical scale: Analytical scale is a type of rating scale that requires teachers to allot separate ratings for the different components of language ability such as content, grammar, vocabulary, etc.

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

19

Assessment: Assessment is a term that refers to a thorough but constant appraisal, judgement and analysis of students' performance through meticulous collection of information. Assessment is a systematic method of obtaining information from tests and other sources, used to draw inferences about characteristics of people, objects, or programmes; the process of gathering, describing, or quantifying information about performance; an exercise such as a written test, portfolio, or experiment that seeks to measure a student's skills or knowledge in a subject area.

Authenticity: Authenticity refers to evaluation based mainly on real-life experiences; students show what they have learned by performing tasks similar to those required in real-life contexts.

Communicative Language Teaching (CLT): In this research, the definition of Communicative Language Teaching (CLT) correlates with that provided by western ELT theorists (e.g. Breen & Candlin, 1980; Ellis, 1990; Savignon, 1991, 2003, 2005; Larsen-Freeman, 1986; Stern, 1992; Brown, 1994; Richards & Rodgers, 2001; Wesche & Skehan, 2002). It refers to a teaching methodology or an approach that focuses primarily on communicative competence comprising both receptive and productive skills (listening, reading, speaking and writing).

Computer-based testing (CBT): Computer-based testing (CBT) is programmed, and then administered to students on computer; question formats are frequently objective, discrete-point items. This type of test is subsequently scored electronically.

Computer-adaptive testing (CAT): Computer-adaptive testing (CAT) presents language items to the learner via computer; subsequent questions on the examination are "adapted" based on a student's response(s) to a previous question(s).

Content validity: When the test accurately reflects the syllabus on which it is based, it can be termed as having content validity. This kind of validity depends on a careful analysis of the language being tested and of the particular course objectives.

Construct validity: Construct validity is the degree to which an instrument measures the construct it was designed to measure; how well an instrument can be interpreted as a meaningful measure of some characteristic or quality.

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

20

Cornerstones of good testing practice: Cornerstones of good testing practice are the guidelines of effective test writers. They include the concepts of validity, reliability, practicality, transparency, authenticity, security and washback.

Criterion-referenced Tests: Criterion-referenced tests are often referred to as standards-referenced tests or proficiency tests. These tests measure how well a student measures up to a certain criterion or standard. Scores tell the test taker how close he or she is to meeting the standard in a given subject.

Curriculum: A curriculum refers to a formal course of study. It is a focus on study consisting of various courses all designed to reach a particular proficiency or qualification. Curriculum is designed to prepare a student for the rigors of a study. The term "curriculum" in this study is seen to include "the entire teaching/learning process, including materials, equipment, examinations, and the training of teachers.

Descriptive statistics: Descriptive statistics describe the population taking the test. The most common descriptive statistics include mean, mode, medium, standard deviation and range; they are also known as the measures of central tendency.

Discrete-point test: A discrete-point test is an objective test that measures students' ability to answer questions on a particular aspect of language. Discrete-point items are very popular with teachers because they are quick to write and easy to score.

Diagnostic test: Diagnostic test is a type of formative evaluation that attempts to diagnose students' strengths and weaknesses vis- a -vis the course materials. Students receive no grades on diagnostic instruments.

Evaluation: Evaluation is described as an overall but regular judgment and analysis of teaching, learning, as well as curriculum through systematic collection of data. Assessment looks at the individual language learners, but evaluation checks the whole language-learning programme. In assessment data is collected by concentrating on students' moment-by-moment performance in the classrooms, "emanating from alternative activities" (Genesee, 2001, p.149) while evaluation involves the gathering of data by focusing on teaching performance and learning outcomes.

High-stakes tests: A high-stakes test is one of such quantitative measures that "occasionally pepper" the subjectivity of school organizations to generate objectivity in education. It has four interrelated components: (1) goals, (2) measures, (3) targets, and (4) incentives (Hamilton et al., 2002).

Holistic scoring: Holistic scoring is based on an impressionistic method of scoring. An example of this is the scoring used with the TOEFL of Written English (TWE).

Face validity: Face validity refers to the overall appearance of the test. It is the extent to which a test appeals to test takers.

Feedback: Feedback helps students reflect on the process of learning as well as the product of that process, and provides specific comments on and specific suggestions for improvement, and encourages students to focus their attention on understanding the task rather than producing a product.

Formative Assessment: Formative assessment is assessment that provides feedback into an on-going academic programme to be used to modify the programme to improve student learning. Assessment is formative when the evidence of learning is actually used to adapt to learning to meet the needs of students, or by students themselves to change the way they work at their own learning. Formative assessment improves learning.

Integrative testing: Integrative testing goes beyond discrete-point test items and contextualized language ability.

Inter-rater reliability: Inter-rater reliability attempts to standardize the consistency of marks between raters. It is established through rater training and calibration.

Item Analysis: Item analysis is a procedure whereby test items and distractor are examined based on the level of difficulty of the item; and the extent to which they discriminate between high-achieving and low-achieving students. Results of item analyses are used in the upkeep and revision of item banks.

Likert Scale: It is a semantic deferential scale that requires subjects to respond to the statements by using a numerical indication of the strength of their feeling towards the object or position described in the statement.

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

22

Low-stakes schools: Low – stakes school refers to a school with high-test scores each year.

Mean: Mean is known as the arithmetic average. To obtain mean, scores are added together, and then divided by the number of students who took the test. The mean is a descriptive statistic. In the present study, mean score is expressed as M.

Mode: Mode is the most frequently received score in a distribution.

Norm-referenced tests: A norm-referenced test indicates how the pupil's performance compares with that of other pupils in some appropriate reference group. A test is considered norm-referenced if the test scores are compared with the scores of a "norming group," which is a representative cross-section of all those taking the test, for example, all eighth-graders taking an eighth-grade math test.

Objective test: An objective test can be scored solely on the basis of an answer key. It requires no expert judgment on the part of the scorer.

Outcomes-based assessment: Outcomes-based assessment focuses on what the student knows and can show. Students compare the outcomes with their learning goals and reflect on the processes that might be changed so that more learning results.

Performance-based test: A performance-based test requires students to show what they can do with the language as opposed to what they know about the language. They are often referred to be task-based tests.

Piloting: Piloting is a common practice among language testers. Piloting is a practice whereby an item or a format is administered to a small random or representative selection of the population to be tested. Information from piloting is commonly used to revise items and improve them. It is also known as field-testing.

Portfolio assessment: Portfolio assessment is one type of alternative assessment. They are a representative collection of a student's work throughout an extended period of time. The aim is to document the student's progress in language learning via the completion of such tasks as reports, projects, artwork and essays.

Practicality: Practicality is one of the cornerstones of good testing practice. It refers to the practical issues teachers and administrators must keep in mind when developing and administering tests, such as time, and available resources.

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

23

Proficiency test: A proficiency test is not specific to a particular curriculum, and it assesses a student's general ability level in the language as compared to all other students who study that language. An example of the proficiency test is the TOEFL.

Range: Range is one of the descriptive statistics or measures of central tendency. The range or min/max is the lowest and highest score in a distribution.

Rating scales: Rating scales are instruments that are used for the evaluation of writing and speaking. They are either analytical or holistic.

Reliability: Reliability is one of the cornerstones of good testing practice. It refers to the consistency of examination results over repeated administrations.

Rubric: Used in the context of assessment, rubric (often scoring rubric) refers to a scoring guide for some demonstration of student learning. It comes from Latin *rubrica* meaning red earth and Middle English *rubrike* red ocher, heading in red letters of part of a book. It is a set of scoring guidelines (criteria) for assessment work and for giving feedback.

Self-assessment: Self-assessment asks students to judge their own ability level in a language. It is a type of alternative assessment.

Standard Deviation: Standard Deviation is a generally used measurement of variability or diversity used in statistics and probability theory. It shows how much variation or "dispersion" there is from the average (mean or expected value). A low standard deviation indicates that the data points tend to be very close to the mean, whereas high standard deviation indicates that the data are spread out over a large range of values. In the present study, standard deviation is expressed as STDV.

Standardized test: A standardized test measures language ability against a norm or standard. It is a test that is constructed in accord with detailed specifications, one for which the items are selected after tryout for appropriateness in difficulty and discriminating power, one which is accompanied by a manual giving definite directions for uniform administration and scoring, and one which is provided with relevant and dependable norms for score interpretation.

Subjective test: A subjective test requires knowledge of the content area being tested. It frequently depends on impression and opinion at the time of the scoring.

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

24

Syllabus: A syllabus is simply an outline and time line of a particular course. It typically gives a brief overview of the course objectives, course expectations, a list of reading assignments, and examination dates. The purpose of the syllabus is to allow the student to work their schedule for their own maximum efficiency and effectiveness.

Institutional curricula and syllabi, generally seen as indispensable units of second/foreign language programmes, can take various forms, can represent various theories of learning, and can be realised in various ways. It is necessary to address confusion in the literature between the terms 'curriculum' and 'syllabus', since these can at times be very close in meaning, depending on the context in which they are used (Nunan 1988, p.3).

In the present study, the HSC English syllabus and curriculum are used as a "single" term because the HSC English syllabus corresponds and represents the HSC English curriculum, and they are inseparable in the Bangladesh context.

Summative test: A summative test refers to a test that is given at the end of a course. The aim of summative evaluation is to give the student a grade that represents his/her mastery of the course content.

Teachers' beliefs: The term here refers to teachers' pedagogic beliefs (Borg 2001), which are related to convictions about language and the teaching and learning of it. These beliefs are manifested in teachers' teaching approaches, selection of materials, activities, judgments, assessment and behaviours in the classroom.

Testing: Test and assessment are both forms of measuring student's language learning ability, but differ in many respects. Tests refer to specific instruments that measure the achievement and proficiency of students whereas assessment refers to a more general concept of scrutinizing students' learning progress.

Validity: Validity is one of the cornerstones of good testing practice. It refers to the degree to which a test measures what it is supposed to measure.

Washback: Washback refers to the influence of testing on the curriculum, teaching, learning, etc. For the purposes of this study, the definition of washback offered by Cheng, Watanabe, and Curtis (2004) in the preface to their book *Washback in Language Testing*, served as the foundation. They state, "Washback (…) refers to the influence of language testing on teaching and learning" (Ibid. p. xiii).

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

25

1.7 Limitations of the Study

A limitation identifies the potential weaknesses of a study. The present study has some limitations with regard to available relevant data in the Bangladesh context. Since no intensive study in this particular area had been carried out in Bangladesh before, the present study suffered from the lack of necessary guidelines and clues that might help it. Another limitation of the study was that it dealt with only the EFL test at the higher secondary level in Bangladesh. It concentrated on investigating how the washback of the HSC examination worked on teaching and learning English as a Foreign Language (EFL) including the syllabi, teaching materials, teaching methods, contents, tasks and activities, and classroom assessment. Since the respondents involved in the investigation were mainly confined to the English teachers and the HSC level students in Bangladesh, this study had no attempt to investigate washback caused by a different types of examination in a different context.

The number of subjects was limited to students, teachers, some examiners and a few curriculum specialists who voluntarily participated in answering and completing the research instruments. Therefore, the results of this study cannot be generalized to a larger population. Data collected in this study was only adequate for describing perceptions of the washback effect of the HSC examination on EFL teaching and learning. Thus, the results would be inappropriate to be generalized to other contexts or other examinations and for other subject areas. Moreover, as the findings in the conclusion were based on the respondents' opinions, further empirical data (e.g., classroom observations), especially from longitudinal studies, should eventually be collected and analysed to add up insight into the nature of this phenomenon, i.e., the HSC examination washback.

Shohamy et al. (1996) reported that washback can evolve over time, so a longitudinal study would perhaps be better able to capture and monitor the ebb and flow of the test impact. However, this would have been impossible in view of the time limitations associated with the current research. Even so, the findings would have been more interesting if the same number of students from different study years had been involved in the study. As it was, only HSC second year students were able to participate. The data for this study were collected over just in nearly 10 months, so follow-up studies are indispensable for observance of long-term washback.

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

26

1.8 Structure of the Thesis

The thesis is constituted of six chapters as follows:

Chapter One "Introduction" offers an introduction to the context of the whole study by giving a brief account of the underlying problems that generated this research study. The chapter incorporates a number of issues, and identifies various components of the problem to be studied including the background information on the general context of the research, relationship between testing and teaching, statement of the problem, objectives of the study, significance of the study, research questions, definition of terms, limitations of the study, structure of the thesis, and a conclusion of this chapter.

Chapter Two "Washback of Public Examinations: Theoretical Framework" incorporates concepts and definitions of washback, its background and origin, its influences on teaching and learning, its connection to impact, its positive and negative connotations, strategies of promoting positive washback and avoiding negative washback, and possible models of the washback process.

Chapter Three "Literature Review" sets out to review of related literature that provides with background knowledge and research insight. This chapter reviews and summarises a good number of washback related research studies with emphasis on the washback effects of tests.

Chapter Four "Research Methodology" deals with the research methods used in the study. It delineates the sampling procedure, development of instruments including the procedures for validating the instruments and building reliability of the instruments, data collection procedures, and analysis of data. This chapter also describes the observation schedule used for classroom observation.

Chapter Five "Presentation and Discussion of the Findings" presents the findings yielded by document analysis, informant interviews, classroom observations, and survey questionnaires. It includes the discussion and interpretation of the findings by synthesizing, integrating, and triangulating the results from different data sets. The findings in this chapter are organised and outlined by themes and patterns.

Chapter Six "Conclusion" is the last chapter of this dissertation. It summarises the findings, answers the research questions, provides theoretical

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

27

implications for the study, proposes a washback model, and suggests some possible directions and recommendations for future research. This chapter presents the findings from both research instruments organised into the pattern established in the previous chapter. It reviews the whole study bringing together themes and results from the earlier chapters. It revisits the concept of washback of the HSC examination in light of the findings of the research. The final section draws conclusion of the thesis based on the findings of the study.

1.9 Conclusion

'Teaching to the test' has become one of the biggest indictments facing the education system at present. It has always been heresy to educators and linguists. Teaching to the test puts too much emphasis on standardized tests that are poorly constructed and largely irrelevant. It stifles creativity and encourages cheating. But today, a new perspective is emerging; it is called curriculum alignment, and means teaching knowledge and skills that are assessed by tests designed largely around academic standards set by the country. Although educators frequently claim that they do not want to teach to a test, the reality is that every educator wants his/her students to be successful with quantitatively high scores. Decision makers, teachers and students equate this success in large part with high-test scores, resulting in classroom instruction that is reflective of test practices and/or expectations.

In this chapter, first, the general context and research problem of this study have been explained. Following that, the significance of and the rationale for the study are presented; the objectives of the study are articulated and some terms are clarified. Then, the organization of the thesis is outlined and the research questions are focused. The following chapter (Chapter Two) clarifies the basic concepts, and explores the theoretical and methodological advances pertaining to washback research. An extensive discussion of the studies in other research areas that influence and shape the present study is also included in the next chapter.

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

28

Chapter Two

Washback of Public Examinations: Theoretical Framework

This chapter focuses on the theoretical underpinnings that shaped and guided this study. It begins with an exploration of the concept of washback by discussing various terms that have been used to describe this educational phenomenon. It then illustrates the mechanism of washback followed by a discussion of the washback phenomenon in different educational contexts such as teaching, learning, syllabus and curriculum, materials, etc. The sections explore how and why washback works to influence other components within the language educational system, trace the rationale behind the use of tests, and examine their power to change teaching and learning. The chapter also presents a review of washback models of teaching and learning in the context of the theoretical and practical considerations of washback. This chapter will guide the research in designing the upcoming chapters.

2.1 Public Examinations: Definitions and Concepts

As defined in the first chapter, public examinations are synonymous with external tests which are administered and scored by external agencies or forces to evaluate learning outcomes or results with a decisive consequence or influence on test-takers. Public exams are often used as instruments to select students as well as a means to control a school system, especially when the educational system is driven by tests or exams (Cheng and Falvey, 2000; Herman, 1992; Smith et al, 1991). That is, public examinations are commonly believed to have an impact on teaching and learning. Given that external tests or public examinations have exerted an influence on teachers and students with an associated impact on what happens in classrooms, such a phenomenon is denoted as washback or backwash.

The origin of public examinations is to be found in the school entrance and civil service examinations of China, which go back at least to the period of the Sui emperors (589-618) (with a prehistory going back much further) and which achieved their most complex form towards the end of the Ch'ing dynasty (1644-1911) (Miyazaki, 1976). Inspired by the Chinese systems, examinations in written format

began to appear in European schools in the 16th century, though it was not until some two hundred years later that public examinations of the type found in China were instituted in Europe for selection to universities, the civil service, and the professions. Public examinations are now a major feature of the educational systems of most European countries, which, in turn, passed them on to their former colonies in Africa, Asia, and the Caribbean, where they still flourish (Kellaghan 1992). The United States, with some exceptions (e.g. the Regents' examinations in New York), has so far not adopted a public examination system. However, during the 1980s and 1990s, a number of proposals contained in reform reports, policy statements, and legislation have advocated a national system or systems of examinations for the country (Madaus & Kellaghan, 1991).

Although there is considerable variation in the form and administration of examinations from country to country (Madaus & Kellaghan 1991; Noah & Eckstein, 1992), they generally share a number of characteristics (Kellaghan 1993). First, the examinations are controlled to varying degrees at national or regional level (and sometimes also administered) by an agency or agencies outside the school (i.e. education board), usually a state department of education, an examinations council closely related to the state department, or regional examining boards. Second, the examinations are geared to syllabi which are usually defined by an agency outside the school, sometimes the same agency as administers the examinations. Third, examinations are usually provided in the traditional areas of the curriculum (such as science, languages). Fourth, examinations are often formal terminal procedures, taken on fixed days under controlled conditions by all candidates taking the examination in a country or region at the end of a course of study. There is a little teacher involvement in assessing students for public examination certification in developing countries. Fifth, examinations are largely written, very often using the essay format, but sometimes making use of multiple-choice items, either in conjunction with other formats or on their own.

There may also be provision for oral and practical assessments in different countries. Finally, as a result of performance on the examination, the student is awarded a grade or mark in each subject examined. Public examinations normally are intended to serve a number of functions. The most obvious is to assess the competence of students' learning relative to some agreed standards. The results are

then frequently used to discriminate among students with regard to their preferred futures: further education, admission to professional preparation, or employment. While certification is important, particularly for students who are leaving the educational system, there is often a danger of losing sight of this function because of the strong emphasis on selection. Examination results are also often used, formally or informally, to provide evidence of school effectiveness, and schools and teachers may be held accountable for their students' achievements as reflected in examination performance. This use becomes more obvious when results for individual schools are published.

In Bangladesh, examination is the only method used for educational measurement. The British Administration imported the public examination here from England. Final External Examination, named Entrance Examination, was started in British India. It was conducted under the rules and regulations of the London University. A student could appear in this examination after completing high school education. This public examination was fit for getting a job under the British Administration. Afterwards, in 1857, the management and controls of this examination was handed over to three universities, i.e., University of Calcutta, University of Bombay, and University of Madras. The system got its full momentum under the Calcutta University up to 1947. Subsequently, it was entrusted to the East Bengal Secondary Education Board at Dhaka and the Dhaka University in respect of the SSC and the HSC examination respectively which were earlier called the Matriculation and Intermediate Examinations.

In 1961, as per the National Education Commission, the then Government of Bangladesh transferred the management and controlling of these two examinations to the Board of Intermediate and Secondary Education of East Bangladesh, from the Dhaka University. The number of education boards was increased in 1963 to cope with the increase in number of schools, colleges and students. At present, ten general education boards, one Madrasha Board, and one Technical Education Board are conducting the public examinations up to XII Grade. Graduate and Postgraduate levels public examinations are being conducted by the National University. Other public and private universities offer undergraduate and postgraduate courses, and conduct examination under the rules and regulations set by the government. The Boards of Intermediate and Secondary Education conduct these examinations.

Students of public, most private, schools sit for these exams. There also exist a different system of public examination at parallel grade levels run by Cambridge International Examinations (CIE) and Edexcel International London Examinations for O' Level (Ordinary level) and A' Level (Advanced level).

In these days, with the widespread adoption of communicative language teaching (CLT) principles, language tests tend to include more practical tasks predicting the real-world settings. During the 1960s and 1970s, language testing techniques were heavily influenced by structural linguistics (Chew, 2005). The analysis of language favoured by behaviourist approaches (e.g. Skinner) led to discrete point testing, that is to say, tests were designed to assess learners' mastery of different areas of the linguistic system in isolation [e.g. grammatical knowledge, vocabulary, pronunciation etc.). Although language testing has been influenced by social changes, there are certain fundamental aspects which remain widely accepted.

Language tests have been used to measure students' achievement for many years. The first book-length discussion of testing English as a foreign language was found in Robert Lado's Language Testing in 1961. Language tests from the distant past to the present are important historical documents. They can help inform researchers about attitudes to language, language testing and language teaching when little alternative evidence of what went on in the bygone language classroom remains. In recent years, there has been a trend towards improving subject matter teaching through the implementation of examinations, especially those defined as "high-stakes" assessment. These efforts are usually part of attempts to introduce changes into the educational system by putting novel pedagogical theories and practices in place; they are related to educational innovation and contribute to building theories on how such innovation occurs. Spolsky (1975) identifies three periods of language testing: the pre-scientific, the sychometric-structuralist and the psycholinguistic-sociolinguistic.

Traditionally, most language tests aim at testing knowledge about the language, such as testing knowledge about vocabulary and grammar. However, according to Brown (2003), "By the mid-1980s, the language testing field had begun to focus on designing communicative language-testing tasks" (p. 10). This means that the need for communicative language test has been recognized, and much research on communicative language tests has been done since then. It was

Chomsky (1965) who first rejected such approaches and proposed an underlying rule-based knowledge system.

From the early 1970s, however, communicative theories were widely adopted among linguistics and they began to focus on "communicative proficiency rather than on mere mastery of structures" in language teaching (Richards 2001, p.153). This trend significantly influenced the methods of language teaching and roles of language testing, although it is highly possible to assume that some social changes induced new theories at first, and then the theories might be modified to support practice more closely. Hymes takes Chomsky's work further, but also reacts against some aspects of it. For Hymes (1972), the social context of language is considered essential and appropriateness was viewed as important as grammatical correctness. Discrete-point teaching and testing models were gradually replaced by models which aimed to integrate the various elements of language learning. A theory of communicative competence has been developed further by Canale and Swain (1980). They also raise two controversial issues related to second language teaching and testing which is explored later:

1. whether communicative competence and linguistic competence are mutually inclusive or separate,

2. whether one can usually distinguish between communicative competence and performance (Spolsky 1985, p.183)

According to the new trends mentioned above, since the 1970s language testers have been seeking more pragmatic and integrative questions for assessment, such as cloze tests and dictations. McNamara (2000) points out the need by stating that the necessity of assessing the practical language skills of foreign students led to a demand for language tests which involved an integrated performance on the part of the language user. The discrete point tradition of testing was seen as focusing too exclusively on knowledge of the formal linguistic system for its own sake rather than the way such knowledge is used to achieve communication.

2.2 Washback: Background and Origin

Washback is a new but very complex phenomenon in the field of education research. It is rarely found in the dictionaries published before 1990s. However, the word 'backwash' can be found in certain dictionaries and is defined as "the unwelcome repercussions of some social action" by the New Webster's Dictionaries, and "unpleasant after-effects of an event or situation" by the Collins Cobuild Dictionary. However, before 1982, no washback study can be traced out either in the field of general education or in the applied linguistics. Washback or backwash, as it is sometimes called, is now a term that is commonly used in the assessment in applied linguistics literature. Although washback is a relatively common term in our field, it is rarely found in dictionaries (Cheng & Curtis, 2004). Because of the importance of the study of Alderson and Wall (1993), as a landmark and milestone in the field of washback research, their study may be considered as an unavoidable work in the washback history.

Kellaghan et al. (1982) are the first who used the term in their work, "The effects of standardized testing" which has extensive potentials for the future researchers. After the work of Kellaghan et al. (1982), other researchers have taken interest to study test washback and to examine how it works on teaching and learning. Between 1980 and 1990, very little empirical research has been carried out to investigate the washback effect of examinations either in the field of general education or in the field of language education.   The other earlier studies in this area are those carried out by Wesdorp (1982) and Hughes (1988).

It should be pointed out that the former (Kellaghan, et al., 1982) was a general education study and not specific to language education. In their ensuing discussion, it is clear that evidence of either beneficial or harmful was often tenuous remaining unproven or, at best, inconclusive. For example, the study of Kellaghan et al. (1982) looks at the impact of introducing standardised tests in Irish schools as a case in point. As early as 1984, Frederiksen publishes a paper called "The Real Test Bias", in which he suggests that because test information is important in attempting to hold schools accountable, the influence of tests on what is taught is potentially great (Gipps, 1994). Nearly 20 years ago, Alderson (1986) identified washback as a distinct and emerging area within the field of language testing. Around the same time, earlier to Alderson, Davies (1985) asked whether tests should necessarily

follow the curriculum. He suggested that perhaps tests ought to lead and influence the curriculum.

Although Alderson (1986) first recognises the potential use of language tests as a tool to bring about positive effects on language teaching and learning about two decades ago, it took almost another 10 years for the concept of tests influencing teaching and learning to become an established research topic. McNamara (2000) argues that this is because applied linguists tend to focus heavily on investigating individuals' language skills and abilities, rather than on the consequences of tests. Wigglesworth and Elder (1996) also point out that the concept of tests influencing teaching and learning is under-researched probably because the huge number of variables involved have made it very difficult for researchers to identify a causal relationship between the test and what went on in the classroom.

Afterwards, Washback on learners was a topic seldom discussed in 1990s, and has gotten more attention from the researchers since the 21st century. The Sri Lankan Impact Study, the first empirical research on washback conducted by Alderson and Wall (1993) is often cited as a landmark study in the investigation of washback. They conducted a two-year investigation of the effects of the implementation of the revised O-Level English examination in Sri Lanka on teaching methodology. The revision of the examination was made to reinforce the innovations in textbooks and teacher training, which were intended to promote communicative English language teaching with its emphasis on practical speaking, reading and writing skills, while discouraging traditional teacher-dominant, grammar focused lessons. The observations of English lessons in 14 secondary schools before and after the implementation of the revised examination revealed that language learning activities and the design of classroom tests were influenced by the new textbooks or tests. However, Alderson and Wall (1993) found that there was basically no difference in the way the teachers taught over the two years of the study as the English lessons remained teacher-centred with little chance for the students to use English in a practical way. They concluded that the positive and desired washback effects were much more limited than expected.

Much of the literature on this subject has been speculative rather than empirically based. The first scholars to suggest that the washback effects of language tests were not as straightforward as had been assumed by Alderson and

Wall (1993). It was Alderson and Wall who pointed out the problematic nature of the concept of washback and the need for carefully designed research. In their article "Does Washback Exist?", they questioned existing notions of washback and proposed a series of washback hypotheses. Within this article they identified 15 hypotheses which may potentially play a role in the washback effect, and must therefore be considered in any investigation (1993, p. 120-121).

Since the publication of the seminal work of Alderson and Wall in 1993, a number of researchers have sought to obtain evidence as to whether washback exists by means of empirical research in language classrooms. With regard to length and duration, the washback studies can be classified in two broad terms: the first kinds have been by definition longitudinal in nature, since they have required the collection of data over a period of time – perhaps two or three school years in the case of revisions to secondary school examinations; and by contrast, studies of the second type have been cross-sectional involving comparisons of teachers, classes, courses and/or schools over a short period of time. Let us look at each kind of research in turn.

2.3 Washback: Definition and Scope

The term 'washback' is commonly used in applied linguistics to refer to the influence of language testing on teaching and learning. In the literature (both in applied linguistics and in general education), the terms 'backwash' and 'washback', are used, and invariably considered as interchangeable. The way standardized tests affect teaching and learning is usually called backwash in educational arena and washback in Applied Linguistics (Karabulut, Aliye, 2007). It has long been affirmed that tests exert a powerful influence on language learners who are preparing to take examinations, and on the teachers who try to help them prepare. It is common to claim the existence of washback (the impact of a test on teaching) and to declare that tests can be powerful determiners, both positively and negatively, of what happens in classrooms (Alderson and Wall, 1993, p. 41). The various influences of tests are often referred to as washback (or backwash). "Washback is the power of examinations over what takes place in the classroom" (Alderson and Wall, 1993, p. 115). Swain (1985, p. 43) succinctly states the prevailing opinion: "It has frequently been noted that teachers will teach to a test: that is, if they know the content of a test

and/or the format of a test, they will teach their students accordingly". Washback can have an individual (micro-level) impact and a social (macro-level) impact. It involves actions and perceptions, influences learners and programmes.

Washback or backwash has been defined as a part of the impact a test may have on learners and teachers, on educational systems in general, and on society at large (Hughes, 2003; Biggs, 1995, 1996; Cheng, 2004). It can generally be understood as the effect of an examination on teaching and learning (Cheng, 2003, Chen, 2002, Hughes, 2003), but all scholars have not agreed to its definition. Alderson and Wall (1993) restrict the use of the term 'washback' to "classroom behaviors of teachers and learners rather than the nature of printed and other pedagogic material" (p. 118). They also consider washback as what teachers and learners do that "they would not necessarily otherwise do" (p. 117). Messick (1996) states that in order to be considered washback, good or bad teaching has to be "evidentially linked to the introduction and use of the test" (p. 16).

Moreover, Wall (1997 in Cheng and Curtis, 2004) makes a clear distinction between washback and test impact. The latter refers to the effect of a test on "individuals, policies or practices, within the classroom, the school, the educational system or society as a whole. Other researchers (Andrews et al., 2002) do not make that distinction, and consider that narrow and wider effects can be included under the term washback. For the purposes of this study, washback was understood in the wider sense including what some scholars call 'impact'. Although being universally used for various purposes, testing is considered by scholars and researchers to induce mostly detrimental washback on teaching.

Tests are often perceived as exerting a conservative force which impedes progress. Andrews and Fullilove point out, "Not only have many tests failed to change, but they have continued to exert a powerful negative washback effect on teaching" (Andrews and Fullilove, 1994, p. 57). These authors also note that "educationalists often decry the 'negative' washback effects of examinations and regard washback as an impediment to educational reform or 'progressive' innovation in schools" (ibid., p. 59-60). Heyneman (1987) has commented, "It's true that teachers teach to an examination. National officials have three choices with regard to this 'backwash effect': they can fight it, ignore it, or use it" (p. 260).

Pierce (1992) states "the washback effect, sometimes referred to as the systemic validity of a test (p.687). In recent years, washback has become a very hot topic among many linguistic and educational experts, and who admit that washback does exist and plays an importance role in language teaching and learning. "There is a natural tendency for both teachers and students to tailor their classroom activities to the demands of the test, especially when the test is very important to the future of the students, and pass rates are used as a measure of teacher success. This influence of the test on the classroom (referred to as washback by language testers) is, of course, very important; this washback effect can be either beneficial or harmful" (Buck, 1988). Bachman and Palmer (1996) consider washback to be a subset of a test's impact on society, educational systems and individuals. They believe that test impact operates at two levels: the micro level, that is, the effect of the test on individual students and teachers; and the macro level or the impact of the test on society and the educational system.

Cohen (1994) describes washback in terms of" how assessment instruments affect educational practices and beliefs" (p. 41). The problem is that while washback is widely perceived to exit, there is little data to confirm or deny these perceptions. This is neatly summarized by Alderson and Hamp-Lyons (1996) in the rationale for their study of TOEFL preparation classes in the United States: "Much has been written about the influence of testing on teaching; however, little empirical evidence is available to support the assertions of either positive or negative washback" (p. 281). Andrews (1994) concurs: "Although a great deal has been said and written about washback, there is in fact relatively little empirical evidence for its existence" (p. 44). Similarly, Shohamy (1993) acknowledges, "while the connection between testing and learning is commonly made, it is not known whether it really exists and, if it does, what the nature of its effect is"(p. 4).

Brown (2000) defines washback as "the connection between testing and learning" (p. 298). Gates (1995) defines washback simply as "the influence of testing on teaching and learning" (p. 101). Alderson and Wall (1993) define washback as the way that tests are (…) perceived to influence classroom practices, and syllabus and curriculum planning (p.17). The influence of the test on the classroom is washback. This influence of the test on the classroom (referred to as washback by language testers) is, of course, very important; this washback effect

can be either beneficial or harmful. Thus Buck's definition stresses the impact of a test on what teachers and students do in classrooms (p.17). Washback is the extent to which the test influences language teachers and learners to do things that they would not necessarily otherwise do (Messick, 1996). The influence of testing on teaching and learning is referred to as washback (Bailey, 1996). Shohamy (1993) summarises four key definitions that are useful in understanding the washback concept:

1. Washback effect refers to the impact that tests have on teaching and learning.
2. Measurement driven instruction refers to the notion that tests should drive learning.
3. Curriculum alignment focuses on the connection between testing and the teaching   syllabus.
4. Systemic validity implies the integration of tests into the educational system and the need to demonstrate that the introduction of a new test can improve learning (p. 4)

Andrews (1994) sees washback as "an influence on teachers, learners, and parents, with an associated impact on what happens in classrooms"(p. 45). Washback sometimes referred to as backwash. Hughes (1989) states "the effect of testing on teaching and learning is known as backwash" (and this term, as he uses it, is synonymous to washback) (p.1). As can be seen, washback is a very complex notion. It can refer to the effect of an examination in the classroom, but also in the school, in the educational system and also in the society. Bailey (1996) states, "washback is the influence of testing on teaching and learning" (p.5).

Pearson (1988) states "Public examinations influence the attitudes, behaviours, and motivation of teachers, learners, and parents, and because examinations often come at the end of a course, this influence is seen working in a backward direction, hence the term, washback" (p. 7). Cheng (2005) concurs that washback indicates "an intended or unintended (accidental) direction and function of curriculum change on aspects of teaching and learning by means of a change of public examinations" (p.112).

Numerous explanations of the term 'washback' can be found throughout the published research and literature on language testing. One of the most common definitions sees the concept referred to as the influence of testing on teaching and learning. Definitions of washback are nearly as numerous as the people who write

about it. These definitions range from simple and straightforward to very complex. Some take a narrow focus on teachers and learners in classroom settings, while others include reference to tests' influences on educational systems and even on society in general. Some descriptions stress intentionality while others refer to the apparently haphazard and often unpredictable nature of washback. From the above illustrations of the definitions of washback, it can be concluded that washback is a subset of a test's impact on society, educational systems, and individuals.

2.3.1 Longitudinal Studies of Washback

A Longitudinal study uses time as the main variable, and tries to make an in depth study of how a small sample changes and fluctuates over time. The present study is a synchronic/cross-sectional Study by nature. A longitudinal study is a correlational research study that involves observations and collecting data of the same items over long periods of time. The reason for this is that unlike cross-sectional studies, longitudinal studies track the same people, and therefore the differences observed in those people are less likely to be the result of differences. Longitudinal studies of washback have generally monitored the impact of innovations in high stakes examinations in particular societies. In some cases, the innovations are revisions to existing examination papers; in others; the examination reform was more radical. This kind of research design requires the gathering of data before the innovation has been implemented, to act as a baseline for the identification of changes in subsequent years as a result of the new or revised exam. Some of the longitudinal studies are stated below.

Li (1990) conducts a longitudinal study of a secondary school leaving examination administered in China – the Matriculation English Test (MET). This high stakes examination had been introduced to replace an older, less valid and less reliable English test. Her methodology involved analyzing 229 questionnaires completed by teachers and teaching-and-research officers. She also analyzed test results and student writing. Her findings were that there was positive washback from the Matriculation English Test (MET) in three areas: (i) a greater use of imported and teacher designed materials which matched the examination requirements; (ii) more classroom time was given to practising the four skills of listening, speaking

reading and writing instead of phonetics, grammar and vocabulary; and (iii) students showed more interest in after-class learning of English.

Shohamy (1993) reported on three longitudinal washback studies she conducted concerning the implementation of three different language tests in Israel and the impact each had on its respective educational system. The first study involved the introduction of an Arabic test by the Ministry of Education. Her research focused on finding out if the test changed teaching practices or student attitudes, and also if there was a long-term impact on teaching. She reviewed teaching materials, interviewed teachers and analyzed student questionnaires and observed lessons. The findings show that there were bigger differences in the initial period of test implementation in terms of materials, class activities, use of mother tongue during teaching and the atmosphere in class. These effects were far less after four years of implementation.

The second study looked at the introduction of an EFL oral test. Shohamy observed and interviewed fifteen teachers. These teachers were divided into two groups: experienced (five years and more) and novice (three years or less). Her results show that experienced teachers were more likely to teach to the test, basing their teaching of oral language on the test, while novices found that the test permitted them to be more creative with activities. The final study examined the introduction of an L1 reading comprehension test towards which teachers had reacted negatively. Shohamy interviewed teachers and analyzed materials produced after the introduction of the test. She found that new materials tended to resemble the test and more time was spent on reading comprehension across the curriculum. Teachers were bitter about the manner in which implementation of the test had occurred, and bitter because they feared that the system would punish them for poor results.

The overall findings, according to Shohamy, indicate that teaching materials and methods cater to the test, and that teachers who have been in the system longer will tend to use the test as teaching guide and curriculum. Cheng (1997) studied the washback associated with a revision of the English Language examination of the Hong Kong Certificate of Education Examination. The study took place from January 1994 to November 1996 and consisted of three phases: general observation and interviews with participants from decision-making bodies, textbook publishers,

principals, department heads, teacher and learners; large scale surveys of teachers and students that occurred in 1994 and 1995; baseline case studies that consisted of classroom observations of nine teachers followed by main case studies of three teachers. Follow up interviews were also conducted with the three teachers. Although the Hong Kong Examinations Authority intended to create a positive washback effect through the innovation, Cheng's findings indicate that changes occurred mainly at a superficial level: the content of teaching and the materials used changed rapidly but there was not much evidence of fundamental changes in teaching practices and student learning.

Wall (1997, 2005) reported on the results of a four-year project in Sri Lanka. The study looked at the effects of implementing a new curriculum and reinforcing the changes by having a new O-level examination (This is the exit level examination for Sri Lankan secondary schools). Wall and the team of designers collected data at three different stages of implementation: (i) prior to the implementation- information was obtained through analysis of official documents, interviews and questionnaires; (ii) during initial implementation – data collection involved classroom observations, examination results and questionnaires; and (iii) full implementation – classroom observations and group interviews with teachers provided the data for this phase of the study. Wall found that although teachers liked the match between curriculum and testing, many other factors, such as the teachers understanding of the requirements of the new curriculum, lack of resource materials, level of difficulty of the examination vis-à-vis the ability levels of the students and prior teaching practices, hindered implementation of certain aspects of the curriculum.

Turner (2002, 2005, 2008) investigated high-stakes test impact at the classroom level in the province of Quebec. She looked at the implementation of a new ESL speaking exam at the Secondary Five level. She wanted to find out whether (1) the introduction of provincial ESL speaking exam procedures affected teacher beliefs, (ii) the introduction of provincial ESL speaking exam procedures affected teaching practices, (iii) there would be a change or pattern in the relationship between teacher beliefs and behavior over time, and (iv) the introduction of provincial ESL speaking exam procedures affected student beliefs

The methodology involved obtaining baseline evidence and evidence after implementation through interviews and classroom observations. Data collection lasted for six months and happened over three time periods. By triangulating the data from the three periods, she found that there was evidence of predictable washback for individual teachers. This evidence, both on the conceptual (beliefs) and instrumental (behavior or practice) level varied across teachers depending on their initial beliefs and practices.

The teachers in this study did not resist the proposed changes implemented via the speaking exam. Instead, they sought to align the required curriculum with classroom teaching and assessment. However, this could partly be attributed to the fact that some of the teachers had participated in prior efforts to develop a rating scale and therefore felt a certain sense of ownership in the ongoing innovations. As a result, feedback and critique was of a more constructive nature. On the other hand, teachers were selective about the changes they elected to adopt – these were chosen with regard to their own established classroom practices and professional stances or beliefs. This reaction seemed to be part and parcel of their professional repertory. The teachers did, however, have difficulties coping with the different goals of classroom-based assessment as opposed to those of the high-stakes provincial exam. She suggests that the results point to the need for better alignment between assessments and the different purposes they are used for. The study also found that student beliefs were affected by the changes.

These longitudinal studies confirm the complex nature of Alderson and Wall's washback hypotheses (1993), which highlight the variable nature of the effect of tests on the various stakeholders. They showed that in some cases there was evidence that over time tests can have a positive impact on classroom activities and materials (Shohamy, Donitsa-Schmidt and Ferman, 1996). However, the implementation of changes to tests in other contexts showed little or no evidence of pedagogic shift (Cheng, 1999, Cheng and Falvey 2000, and Qi, in press). The benefit of a longitudinal study is that researchers are able to detect developments or changes in the characteristics of the target population. The key point here is that longitudinal studies extend beyond a single moment in time. As results, they can establish sequences of events.

2.3.2 Synchronic/Cross-Sectional Studies of Washback

A cross-sectional study takes a snapshot of a population at a certain time, allowing conclusions about phenomena across a wide population to be drawn. This approach to washback research has involved a focus on existing tests or examinations, using a comparative design. This kind of study is conducted over a relatively short period of time, making it more practical for many researchers than the more extended, longitudinal types. Watanabe conducted several studies on examinations within the Japanese context (1996a, 1996b, 2004a, 2004b). In his first study (Watanabe 1992), he hypothesised that Japanese students who had sat the university entrance examinations would have more restricted learning strategies than those of a control group of students who were able to enter university via a system of recommendation rather than examination.

Andrews (1995) conducted a study on the addition of an oral component to the Hong Kong Use of English Examination (UE). Alderson and Hamp-Lyons (1996) examined TOEFL preparation classes for evidence of washback. They were interested in finding out more about how teachers describe the way they teach to prepare their students for the TOEFL test. Snyder et al. (1997) investigated the experience of Uganda in trying to change teachers' classroom practices by manipulating high-stakes testing. They looked at the extent to which a new version of the Primary Leavers Examination, implemented by the Uganda Ministry of Education, led teachers to change their teaching practices. Watanabe (1996, 2004) conducted a washback study that focused on the high stakes English entrance examinations for Japanese universities.

Greene (2007) looked at preparatory courses in the United Kingdom for students taking the IELTS Academic Writing Component (AWC). He observed two types of classes using a modified version of COLT: he noted and coded activities in classes doing IELTS AWC preparation and regular English for Academic Purposes (EAP) classes in UK institutions for a period of twelve weeks. He did not find a significant washback effect – statistical analyses showed that the classes were essentially similar in terms of time spent on specific writing activities. He postulates that the test design of a high-stakes exam does not have strong washback effect in these institutions; teaching in EAP courses could be influenced more by institutional variables such as the teachers' level of professional training, and teacher factors such

as their beliefs about effective learning. Shih (2007) and Pan (2009) both examined the effects of English exit certification in Taiwan. Shih (2007) focused specifically on the General English Proficiency Test (GEPT) and was particularly interested in the washback effect on learners. The results indicate that this test had various but limited washback effects on the learning of participants.

W hat is clear from these studies is that a test dose not have the same effects on all teachers preparing students to take it. The reasons for this seem to stem from decisions, expectations and assumptions made by all stakeholders from test developers, administrators, materials and syllabus designers, through to teachers and students. The reasons why teachers teach the way they do, and in essence the fact that they are teaching test preparation at all, seems inseparable from the other elements which create the context that they teach within. The benefit of a cross-sectional study design is that it allows researchers to compare many variables at the same time. The present study is a cross-sectional study by nature.

2.4 Types of Washback

Generally, washback can be analysed according to two major types: *positive and negative,* depending on whether it has a *beneficial* or *harmful* impact on educational practices. For example, a test may encourage students to study more or may promote a connection between standards and instruction. Washback from tests can involve individual teachers and students as well as whole classes and programs. Bachman (2000) terms washback as: *macro contexts*, and *micro contexts*. The *micro level*, the effect of the test on individual students and teachers; and the *macro level*, the impact the test may have on society and the educational system.

Some kinds of washback result from the effects of a test on the language learners themselves, while other kinds of washback are more closely related to effects of a test on personnel involved in language teaching (including influences on teachers, administrators, course designers, and materials developers ultimately influencing courses, programs and materials). Bailey (1996) calls two sorts of washback: *learner washback* and *program washback*, respectively. This idea overlaps, to some extent, Bachman and Palmer's (2000) *micro* and *macro levels of washback*, although they have included the influences on individual teachers under the micro category.

2.4.1 Positive Washback

Hughes (1989) suggests, 'If you want to encourage oral ability, then test oral ability' (p. 44). Positive washback is said to result when a testing procedure encourages 'good' teaching practice. For example, an oral proficiency test is introduced in the expectation that it will promote the teaching of speaking skills. Positive washback would result when the testing procedure reflects the skills and abilities that are taught in the course, as, for instance, with the use of an oral interview for a final examination in a course in conversational language use." Therefore, when there is a match between the activities used in learning the language and the activities involved in preparing for the test, we say that our test has positive washback. The following figure shows how a washback works on syllabus and curriculum.

Positive washback can be used to influence the language syllabus and curriculum. As Davies (1990) mentions, washback is inevitable and it is foolish to pretend that washback does not happen. Therefore, in order to prepare students for the examination, the communicative way of teaching will be adopted in our classes and this positive washback helps us change the curriculum the way we want. Positive washback can be summarised as below:

Firstly, teachers and learners will be motivated to fulfill their teaching and learning goals (Anderson & Wall, 1993). Secondly, positive washback takes place when tests induce teachers to cover their subjects more thoroughly, making them complete their syllabi within the prescribed time limits. Thirdly, good tests can be utilized and designed as beneficial teaching-learning activities so as to encourage a positive teaching-learning process (Pearson, 1988). Fourthly, a creative and innovative test can quite advantageously result in a syllabus alteration or a new syllabus (Davies, 1990). Fifthly, examination achieves the goals of teaching and learning, such as the introduction of new textbooks and new curricula (Cheng; 2005). Sixthly, tests induce teachers to cover their subjects more thoroughly, making them complete their syllabi within the prescribed time limits. Seventhly, tests motivate students to work harder to have a sense of accomplishment and thus enhance learning. Eighthly, good tests can be utilized and designed as beneficial teaching learning activities so as to encourage positive teaching-learning processes. Finally, decision makers use the authority power of high-stakes testing to achieve

the goals of teaching and learning, such as the introduction of new textbooks and new curricula.

2.4.2 Negative Washback

Negative washback is said to occur when a test's content or format is based on a narrow definition of language ability, and so constrains the teaching/learning context. If, for example, the skill of writing is tested only by multiple choice items then there is great pressure to practice such items rather than to practice the skill of writing itself'. As Brown (2002) states washback becomes negative washback when there is a mismatch between the content (e.g., the material/ abilities being taught) and the test. Washback is harmful:

a) when training for a particular test comes to dominate classroom work;
b) when teachers teach one thing and the test then concentrates on another one; and
c) when teachers end up "teaching to the test".

Actually, much teaching is always directed towards testing and much time of the class is spent on materials that appear in the test. Sometimes, the objectives and contents of the test do not appeal to students and teachers. For example, some students like and need to learn English communicatively, but the test they have to undergo is discrete-point. Both positive and negative washback work at both level: micro-level (classroom settings), and at macro-level (educational and societal system). Some of the reasons as well as the outcomes of the negative washback are illustrated below:

a) Test comes to dominate classroom work,
b) There is no correlation between test objectives and curriculum objectives.
c) Teachers teach one thing and the test then concentrates on another one,
d) Teachers tend to ignore subjects and activities that are not directly related to passing the exam, and tests accordingly alter the curriculum in a negative way.
e) Students may not be able to learn real-life knowledge, but instead learn discrete points of knowledge that are tested.
f) Tests bring anxiety both to teachers and students and distort their performance.

g) Teachers tend to ignore subjects and activities that are not directly related to passing the exam, and tests accordingly alter the curriculum in a negative way.

h) The tests fail to create a correspondence between the learning principles and/or the course objectives.

i) An increasing number of paid coaching classes are set up to prepare students for exams, but what students learn are test-taking skills rather than language learning activities.

j) Test narrow down the curriculum, and put attention to those skills that are most relevant to testing.

k) Decision makers overwhelmingly use tests to promote their political agendas and to seize influence and control of educational systems.

Likewise, Shohamy (1992) identifies some of the conditions that may lead to negative washback:

a) When reliance is on tests to create change,

b) When emphasis is mostly on proficiency and less means that lead to it,

c) When tests are introduced as authoritative tools, are judgmental, are prescriptive, and dictated from above, and

d) When the writing of tests does not involve those who are expected to carry out the change- the teachers.

The question is how to promote the intended washback of a test and minimise the possible counterproductive reactions. First, the test must accurately reflect course objectives and the principles of mastering the knowledge need. This will lead teachers and learners to appropriate teaching and learning styles and enable beneficial washback to operate. If the test is at variance with the course objectives, it will require teachers to focus their teaching on the test alone and cause harmful washback. Secondly, teachers, administrators and others involved should be trained and provided with information concerning the test, such as the aims, item type, scoring systems, specimen papers, etc.

Competence and familiarity will help teachers and administrators to work properly toward the test, and limit misuse of test and its results (Swain, 1985). Next, test consequences play an important role in enabling either beneficial or harmful

washback to operate. The more profound the consequence, the greater washback effect is. Educational settings would help to balance beneficial and harmful washback in reducing test pressure toward teachers and students by appropriate continuous assessment. Furthermore, apart from the test itself there are many factors within a society, particularly the educational environment with its typical conditions all influence the behaviours of teachers and students. Nevertheless, to what extent these factors operate much depend on how they interact with each other in a specific circumstance.

2.5 The Mechanism of Washback

Washback is not as straightforward as it was previously thought. Its mechanism is complicated. Mechanism of washback refers to how washback works on macro and the micro level, positively and/or negatively. Tests have often been used at the end of the teaching and learning process to provide a diagnosis of the effects of teaching and learning. However, testing may well be considered before the teaching and learning, in order to influence either or both processes. This view of testing is derived from the realisation of test power and its manifestations with regard to high-stakes decisions based on test results for individuals, educational systems and society as a whole. This section looks at the functions and mechanisms by which washback works in relation to other educational theories and practices.

Understanding of washback mechanism can be more deepened by the observing the different models of washback. Unlike the Washback Hypothesis, which only proposes a linear relationship between tests and teaching or learning, Bailey's (1996) model emphasises the importance of the interaction among the different components. Washback variables influencing various aspects of learning and teaching can be divided into "washback to the learner" and "washback to the programme" (Bailey,1996, 1999); the former refers to the impact of the test on test takers, while the latter is concerned with the impact of the test on teachers, administrators, and curriculum developers. The washback effect, however, is not solely confined to teaching and learning. Variables such as materials, curriculum and research are encompassed, making the mechanisms of washback more intricate and comprehensive. The methodologies used in this area have mainly been surveys, interviews and observations. In this respect, Watanabe (2004a) has pointed out, there

are perhaps effects on teaching and learning that interviews and observations alone or combined may not be able to capture.

Cheng (2002, 2004) mentions the importance of considering factors such as a society's goals and values, the educational system itself, as well as approaches to teaching and learning within the system in washback analyses. Watanabe (2004a) and Cheng (2004) both suggest that ethnographic, triangulation methods should be carried out to push the boundaries of what can be discovered about the washback effect. Empirical evidence from these types of data collection efforts should provide stronger, more comprehensive bases on which to theorise washback models. Efforts in this direction have already begun. Over the past two decades, several models have been proposed concerning washback. In the next section, some of the models are presented.

2.5.1 Washback Models

There have been few attempts to describe a model of how a test can influence teaching and learning. This may indicate the difficulty of finding patterns of the way tests influence teachers and students. The impact of an assessment seems to depend not only on the quality of the assessment itself and the way the results are used, but also the context in which the assessment is introduced and administered, and the beliefs held by stakeholders such as teachers and students.

During recent years, though a good number of washback studies have been carried out, the washback models are still to be adequately defined and analysed. In the field of applied linguistics, there seem to have been some attempts to create a model which might illustrate the mechanism of washback. The models of washback discussed below evolve as more research findings became available and a clearer picture of the nature of washback emerged. Thus, the models illustrate the shift in views of washback over the past nearly 20 years. The traditional model of washback emerges in the early 1990s prior to the study by Alderson and Wall (1993). It is characterised as the trichotomy model proposed by Hughes (1993).

Washback models, in general, have been adapted from models or frameworks suggested in language testing, EFL and educational innovation literature. A common characteristic of these washback models is that they tend to

highlight what washback looks like and who is affected, but do little to address the factors that contribute to the phenomenon. In other words, "process" is less understood than "participants" and "products". Besides, the products in these models/hypotheses refer mainly to teaching and learning washback, not to the aspects of washback that might impact society. Some specific models that have been proposed in washback literature, and how these they have been developed, are discussed in this section

2.5.1.1 Hughes's Washback Model

Hughes's (1993) model is a pioneer washback model in applied linguistics. In discussing the complex mechanisms through which washback occurs in actual teaching and learning environments, Hughes (1993) introduces a concept of trichotomy and argues for distinguishing between participants, processes and products in both teaching and learning, recognising that all three may be affected by the nature of a test. In the Hughes's model (Table 2.1), 'participants' are students, teachers, administrators, materials developers and publishers, whose perceptions and attitudes towards their work may be affected by a test. In his unpublished paper cited by Bailey (1996), and Cheng and Curtis (2004), Hughes (1993) made a distinction between participants, process, and products:

Table 2.1: Hughes's trichotomy of backwash model

(a)	Participants – students, classroom teachers, administrators, materials developers and publishers, whose perceptions and attitudes toward their work may be affected by a test
(b)	Processes – any actions taken by the participants which contribute to the process of learning
(c)	Products – what is learned (e.g., facts, skills, etc.) and the quality of the learning (e.g., fluency)

Hughes uses the term 'processes' to cover any actions taken by the participants which might contribute to the process of learning, such as the development of materials, syllabus design, and teaching methods. Finally, 'products' refer to what is learned (facts, skills, etc.) and the quality of the learning (fluency, etc.). The trichotomy into participants, process and product allows planners to

construct a basic model of backwash. Hughes (1993) suggests that the nature of a test may first affect the perceptions and attitudes of the participants towards their teaching and learning tasks. These perceptions and attitudes in turn may affect what the participants do in carrying out their work (process), including practicing the kind of items that are to be found in the test, which may affect the learning outcomes, the product of the work. As a pioneer model, it attempts to clarify how test works to desired outcomes. However, the model does not sufficiently clarify the term 'processes'. As a first model of washback, it received worldwide recognition.

2.5.1.2 Bailey's Washback Model

Based on Hughes' (1993) tripartite distinction between participants, processes and products, Bailey (1996) develops and illustrates a model in which a test not only affects products through the participants and the processes they engage in, but where the participants and processes also in turn provide feedback and thereby also has an impact on the test, as dotted lines in Figure 2.1 represent. This model is an early attempt at theorising washback, but is not empirically grounded.

This model incorporates ideas from Hughes (1993) in describing a trichotomy of test effects in terms of "participants", "process", and "product". Her model, however, is innovative in that it is grounded in empirical research evidence from educational change taking place in the Hong Kong context. Bailey points out participants include students, teachers, materials writers, curriculum designers, and researchers. Here, the participants refer to the stakeholders who directly participate in the teaching, learning, and testing process. Processes refer to the ways teaching is executed. Processes, according to Hughes (1993), refer to material development, syllabus design, changes in teaching methodology, and testing strategies among others. The products in a washback study refer mainly to what are learned and achieved. Products include learning, teaching, new materials and curricula, research results. Here, the focus is the development of communicative competence:

Figure 2.1: Bailey's washback model (1996)
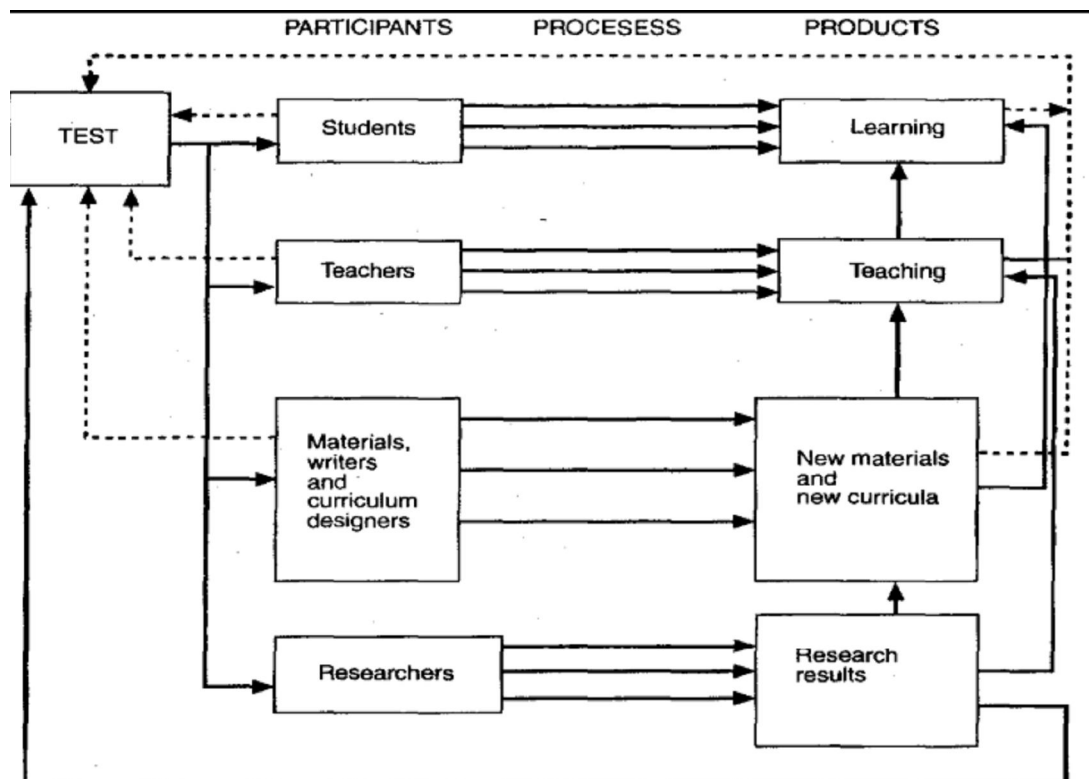
Bailey's model is designed on the basis of suggestions of Hughes (1993); however, she does not clarify the process herself. Bailey's model (Figure 2.1) shows and describes the participants and products, but it does not give any information of process. An apparent shortcoming in this figure was that it showed a test directly influencing the participants, without articulating the role of beliefs held by the participants. In other words, the model did not explain why the participants did what they did. In addition, the model proposed by Bailey (1996) no longer finds strong support among researchers as a model of washback because it includes wider test effects such as those on teaching materials which can be referred to as *impact*, rather than being restricted to the effects that a test has only on teacher and learner behaviour (i.e., washback) as defined by Hamp-Lyons (1997) and Wall (1997). However, her model has immensely contributed to the washback studies during the last decade. Her model can be considered as a gateway and one of the pioneer washback models for future researchers.

2.5.1.3 Burrows's Washback Models

Another set of simpler models is presented by Burrows (1998). She seeks empirical evidence of the washback effect on the attitudes and practices of teachers on the Adult Migrant English Program in New South Wales in Australia (Figure 2.2). Her study looks at the impact of the implementation of the Certificate in Spoken and Written English. Her conclusions are that there is evidence of washback, but that different teachers react to the changes in assessment differently. She also feels that in her case, where testing and the curriculum were closely interwoven, the changes were not easy to separate.

Burrows (1998.) identifies three models of washback: one traditional pre-dating Alderson and Wall (1993); a second model, relating to current writing about washback (e.g. Shohamy et al., 1996); and she proposes a third model relating washback to curriculum innovation and teachers' beliefs, assumptions and knowledge (BAK) as shown in the following diagrams (Figure 2.2):

Figure 2.2:   Burrows's washback models (1998)
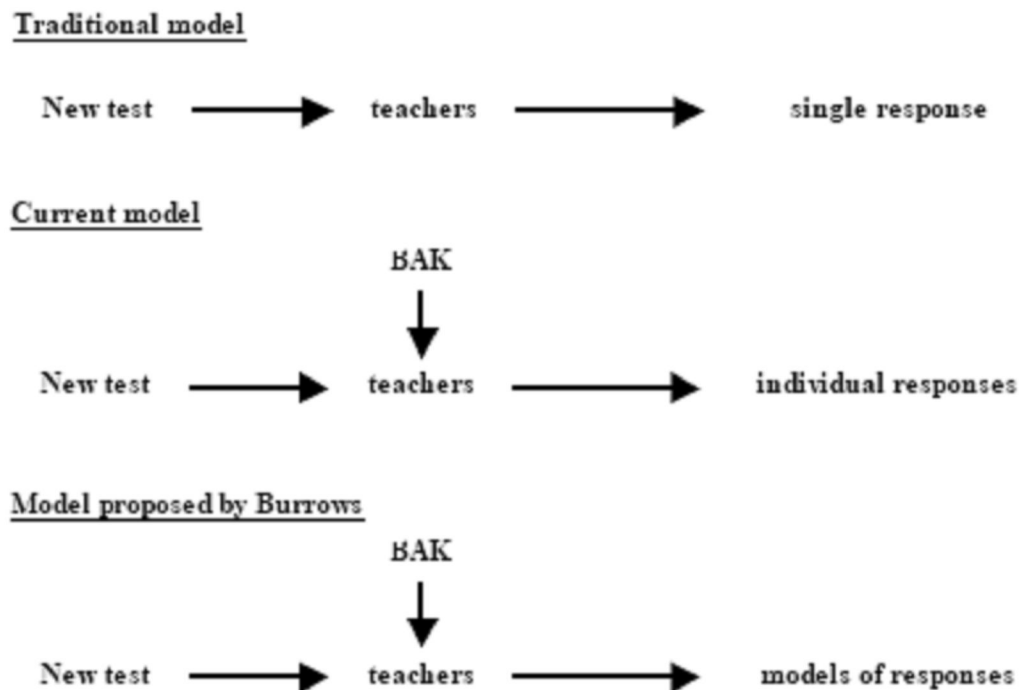
Burrows (ibid.) has argued that the models imply that a uniform and consistent washback effect would always be expected by the introduction of any test because the washback depends on the quality of the test rather than on the participants. She suggests that this early model is not based on objective evidence such as observation, but on teachers' anecdotal evidence. However, Burrows's models lack of discussion on the role of participants and teaching methodology. The models fail to draw wide attention of researchers due to their limitations.

2.5.1.4 Cheng's Washback Models

Cheng (1999) proposes a model of washback and identifies three levels of washback effect of the 1996 Hong Kong Certificate of Education Examination (HKCEE) in terms of curriculum change. Agencies of the three levels are (1) decision-making agencies, (2) the intervening agencies, and (3) the implementing agencies (Figure 2.3). The HKEA makes the decisions with its subject committee, which consists of persons nominated by the Director of Education, English subject examination officer, language experts from tertiary institutions and school teachers. The HKEA piloted the revised syllabus and went to schools to get opinions from teachers:

Figure 2.3: Cheng's explanatory washback model (1999)

Level 1
Decision making Agencies

CDC
Teaching syllabus
HKEA
Exam syllabus

How do the two syllabuses work? How are the teaching and learning principles behind the exam syllabus realised as manageable tasks?

GAP

Level 2
Intervening Agencies

Textbook publishers
Teaching contents
Tertiary institutions
Teaching methods

How do teachers realise the teaching and learning theories behind the exam syllabus and the textbooks? How do teachers arrange their lessons and activities accordingly?

GAP

Level 3
Implementing Agencies

Schools
Principals
Teachers
Students

How are the teaching and learning activities carried out in classrooms?

The model suggests that the textbook publishers revise textbook materials, and also inform tertiary institutions about further teacher education. Cheng (1999) points out that it is up to the schools and teachers to decide how they are going to carry out their teaching according to the syllabus. Such a process usually signifies a cycle of a curriculum change. Cheng (1999), in her model, suggests that teachers and principals redefine and reinterpret the messages about policy that they receive; they then act - adapt, teach, learn, and evaluate - according to their own definitions of the situation . Therefore, the identification of the gaps noted in Figure 2.3 would greatly improve the knowledge and understanding of how and in what areas a public examination change can actually influence the Hong Kong school curriculum. Her model proves to be significant in that she describes three levels of agencies and their responsibilities. It is assumed that her model would be more powerful if she would suggests any teaching methodology and teacher training

Cheng (2002) comes up with another model (Figure 2.4) of washback based on her study of the Hong Kong Certificate of Education Exam (HKCEE). Her model is specifically for the Hong Kong educational context. A diagram of the model she has proposed is shown in figure below (Figure 2.4):

Figure 2.4 Cheng's washback model (2002)

New 1996 HKCEE	Participants	Processes	Products
Level 1 Decision making Level	• ED • HKEA & CDC	syllabus design	new HKCEE
Level 2 Intervening Level	Textbook publishers Teacher educators	*materials development bring about changes in teaching and learning through teacher education*	*new teaching materials*
Level 3 School Level	Principals Department heads Teachers &Students	*Changes in attitudes toward teaching and learning; changes to teaching and learning activities*	
			improved learning out comes

Cheng (2002) has obtained empirical data from a longitudinal study using a mixed methods approach that emphasised the importance of context, setting and subject frames of reference to examine the washback effect of the new Hong Kong Certificate of Education Examination in English HKCEE. She looks at changes to the public examination system in Hong Kong, specifically to the HKCEE and the impact on teacher as well as student behaviours in the classroom. Her model shows that there are levels of participants, processes, and products as delineated by Hughes (2003). In addition, Cheng's (2002) model describes the role of participants. The model also shows the activities under process.

Three major research questions were explored over three phases of this study. These questions are: (1) What strategies did the Hong Kong Examinations Authority (HKEA) use to implement the examination change? (2) What were the nature and the scope of the washback effect on teacher and learner perceptions of aspects of teaching for the new examination? (3) What were the nature and scope of the washback effect on teacher behaviour as a result of the new examination? Cheng's both two models are based on curriculum innovation and language teaching. Her models (Figures 2.3 and 2.4) are praiseworthy and can be modeled for other researchers.

2.5.1.5 Chapman and Snyder's Test Impact Model

Chapman and Snyder (2000) attempt to describe a model how a test can influence teaching and learning. Based on the international educational development literature, Chapman and Snyder (2000) devise the model which illustrates what they call "logical paths" through which policy makers assume that the use of high-stakes assessments may bring about improved student learning.

Four possible uses of tests are shown on the left of the model and the intended outcome (i.e., improved learning) is shown on the right. The paths linking them have intermediate events which may include community pressure, but all include providing extra resources and improvement of instructional practice. All arrows point to assumed direct consequences. They (ibid.) argue that the model is very simplistic as it does not take into account the complexity of the teaching and learning process, or that teaching and learning may not be easily altered just by

manipulating single factors. The figure below (Figure 2.5) presents the model of washback proposed by Chapman and Snyder (ibid.):

Figure 2.5: Chapman and Snyder's test impact model (2000)

Possible use of testing   Linkages to instructional practice
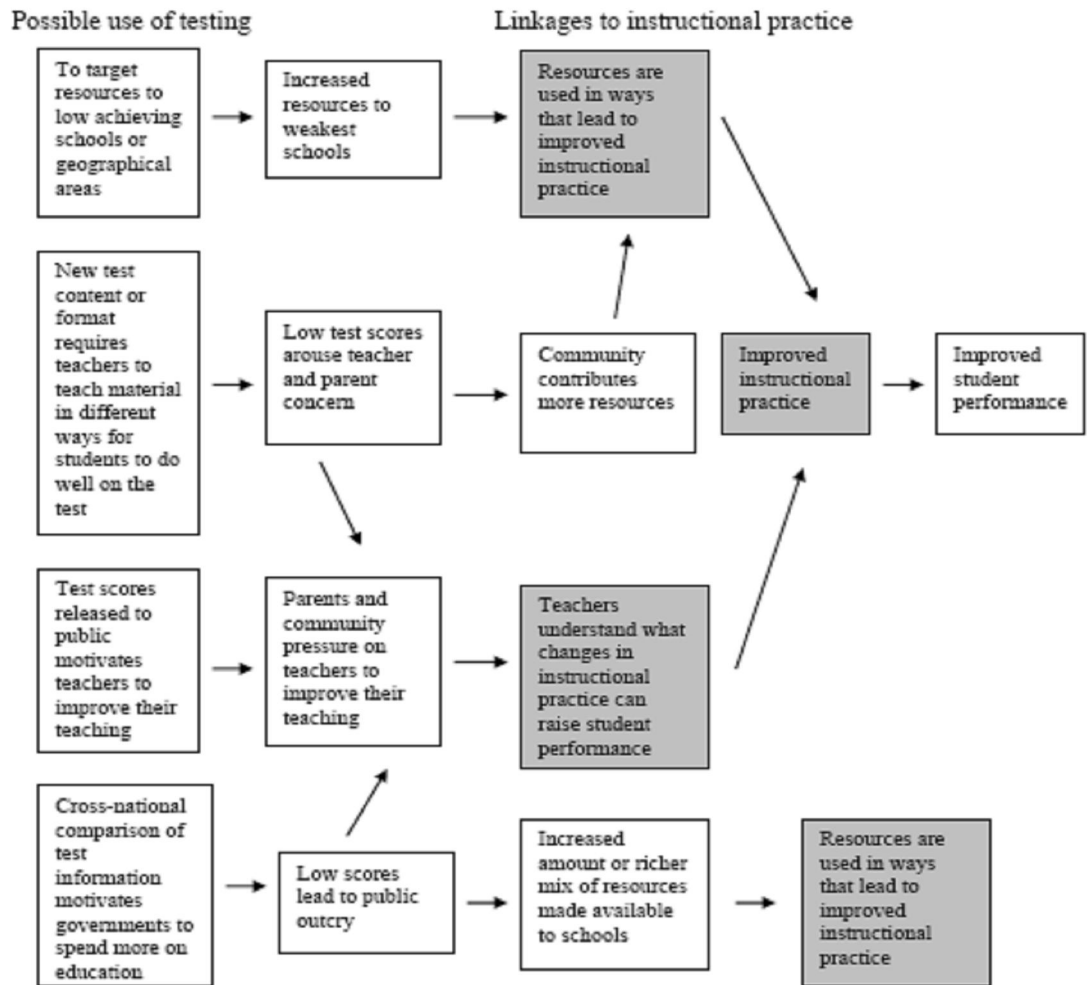
They (ibid., 2000) have stressed that policy makers were responsible for clarifying and elaborating the link between testing and improved teaching and learning. The model discussed above presents teachers as rather passive, as if their beliefs have no part to play in the process. Although They (ibid, 2000) do not articulate the role of beliefs in the model, it can be argued that one of the embedded assumptions is belief change as Fullan (2001) suggested that it would play an important role in promoting desired test impact. The model (Figure 2.5) is a complex and ambitious one. The linkage, they try to establish is hardly possible to happen.

Green's Washback Model

Green (2003) proposes a predictive model of test washback set out in Figure 2.6. In considering the mechanisms of washback, a growing body of theory relates test design, test use, and classroom behaviours. These embrace both contexts for test use and technical qualities of the test instrument. Green (2003) tries to draw together these two elements in washback theory by introducing the model. The model starts from test design characteristics, and relates validity issues of construct representation identified with washback. In the proposed model, test design issues are most closely identified with the direction of washback—whether effects are likely to be judged beneficial or damaging to teaching and learning.

The model below (Figure 2.6) relates design issues to contexts of test use, including the extent to which participants (including material writers, teachers, learners, and course providers) are aware of and are equipped to address the demands of the test and are willing to embrace beliefs about learning embodied therein:

Figure 2.6: Green's washback model (2003)
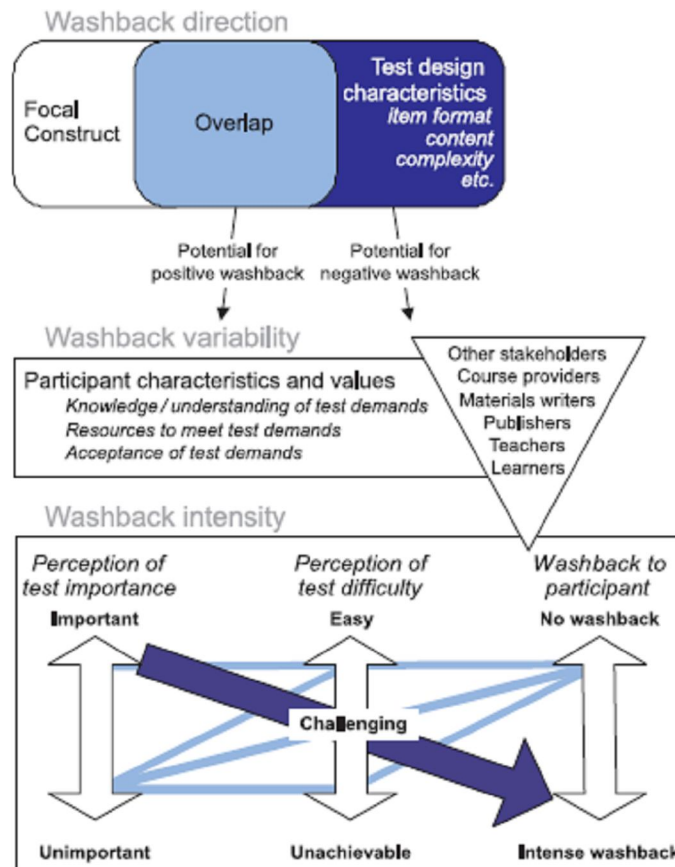
In this model, these features are most closely related to washback variability (differences between participants in how they are affected by a test) and washback intensity. Green (2003) suggests that washback may be most intense—have the most powerful effects on teaching and learning behaviours— where participants see the test as challenging and the results as important (perhaps because they are associated with high stakes decisions, such as university entrance). The model also indicates that the conditions for intense washback to a majority of participants would seem to be in place. The model seems to be very complex because it tries to relate theory, test design, test theory, and classroom behaviours. It is a washback model of direction, variability, and intensity.

2.5.1.7. Manjarrés's Washback Model

Manjarrés (2005) designs a model of washback to show how it works in the context and the type of washback that the different factors seem to be generating. Manjarrés (2005) suggests that test produces general awareness of the importance of English, reduced class size and seems to contribute to the generation of ideal goals in line with the communicative competence construct. These are in themselves part of the general positive washback effect, which was perceived here as 'strong' and 'positive'. The figure below (Figure 2.7) displays how participants, processes, and products are coordinated to promote students' level of communicative competence.

Figure 2.7: Manjarres's washback model (2005)

Value/ Specificity	Factors Mediating Washback	Participants	Processes	Products
General Positive Washback	Macro – context: importance of English	School decisions: number and size of classes	- - -	- - -
	Perception of English in the society: educational legislation	Teacher's awareness		
		Student's awareness		
		Ideal goals		
Specific Negative Washback	Importance of exam in the context	Syllabus Specific objectives Classroom activities and interaction In-class evaluation practices		Students' level of communicative competence
	Teacher's beliefs about how language is learnt and what communicative competence is			

However, since it is 'general' washback, as a factor for the outcome of the test in terms of learning, its effect shows to be rather weak. What seem to be crucial are the teacher and the decisions he/she makes (syllabus, activities, evaluation, etc.). These decisions, however, cannot evidentially be linked to the examination because nothing in the class or in the interviews can uncontroversially show such a direct relation, but there appears to be, nevertheless, a strong correlation. The model depicts different factors meditating washback, but it does not include teacher training and teaching methodology which are very influential factors in generating washback.

2.5.1.8 Nguyen's Washback Models

Nguyen (2005) proposes two models of washback on the teacher-level and student-level. The figure below (Figure 2.8) displays the circle of testing effects on teacher-level. In the model, the double directional arrow from one factor to the other factor indicates the direction of the influence from the determining factor to the dependent one. The other directional arrow shows in turn interaction, the dependent factor becomes the determining one. These interrelationship forms a circle of the causal links:

Figure 2.8: Nguyen's test washback model - effect on teachers (2005)
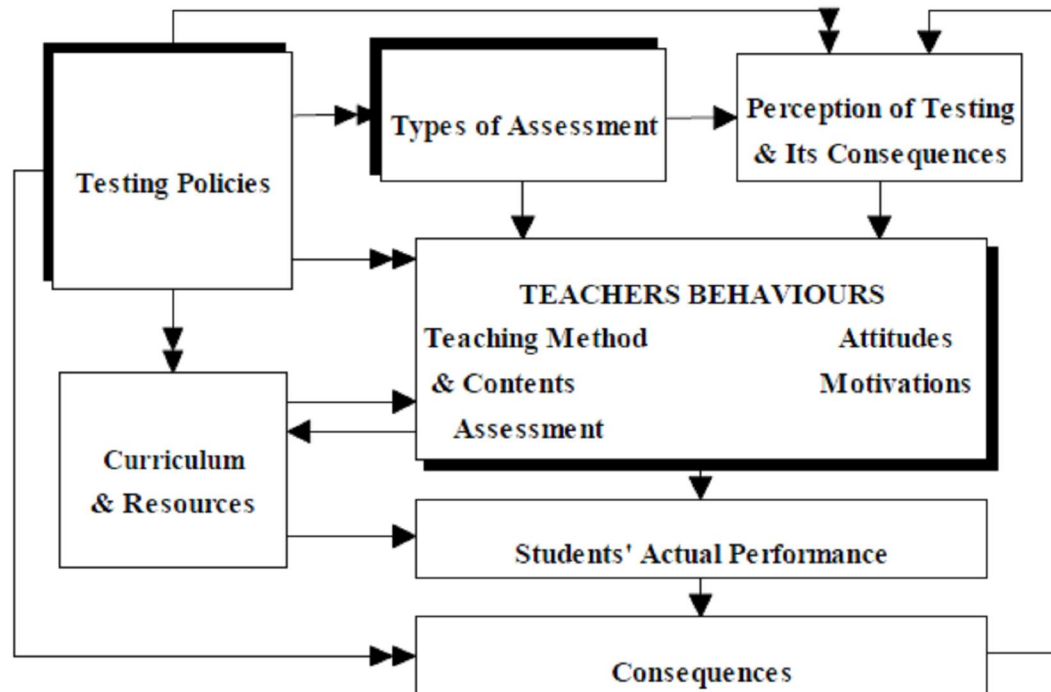
Examining the model (Figure 2.8) from left to right, it is seen that testing policy is the primary determining factor that can be intervened to enable either positive or negative washback on types of assessment, teachers' perception of testing and its' consequences, teachers' behaviours, consequences of the test results and curriculum and resources. Furthermore, next, types of assessment play a very important role that together with the testing policies may influence teachers' perception of testing and test types. They enhance changes in teachers' behaviours that lead to changes in attitudes and motivations and teaching content and method. The model (Figure 2.8) reflects that curriculum, resources, and teachers' behaviours interact with each other in two ways that indicate by two arrows in opposite way. The model suggests that the curriculum and resources also directly influence students' actual performance. The model highlights that the outcomes of the changes and interactions lead to change in students' actual performance then consequences.

Nguyen (2005) also proposes another washback model. In the model below (Figure 2.9), the double directional arrow from one factor to another factor indicates the direction of the influence from the determining factor to the dependent one. The other directional arrow shows in turn interaction, as the dependent factor becomes the determining one. These interrelation forms a circle of the causal links:

Figure 2.9: Nguyen's test washback model - effect on students (2005)
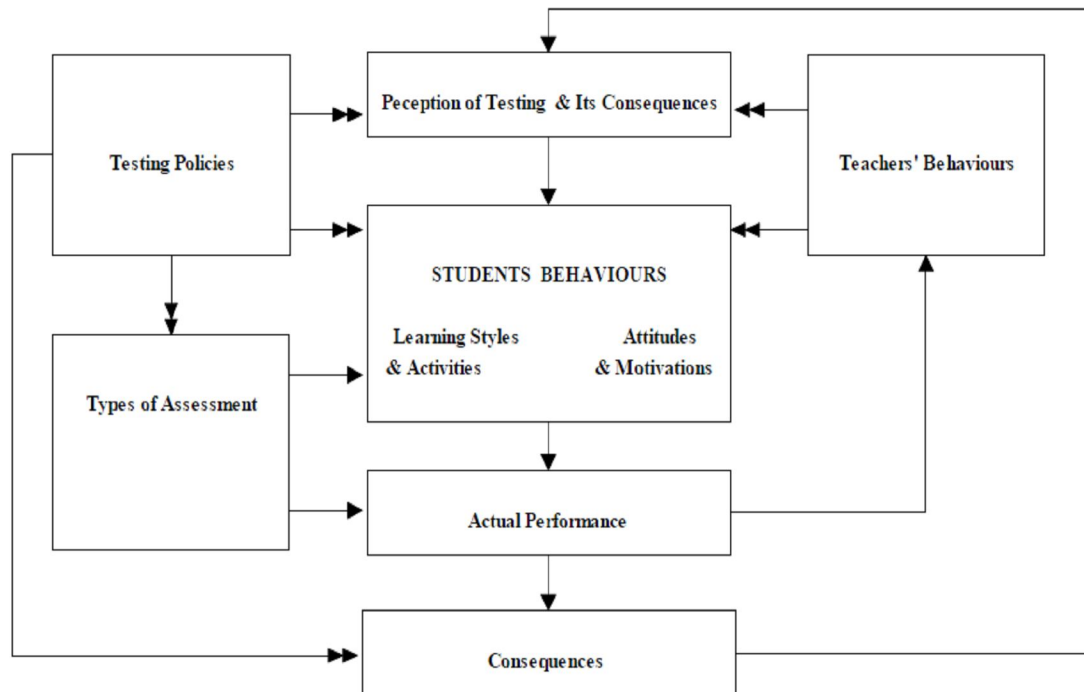
Nguyen (2005) suggests that testing policy is the primary determining factor that influences students' perception of testing and its consequences, types of assessment and the consequences of test results. The two models (Figures 2.8 and 2.9) suggest that test washback effects, or more specifically content and method washback, pressure washback, and innovations in education can primarily be promoted by the testing policies and types of assessments, then teachers' perception of the testing policies and of the type of assessment in use. Hence, to enhance beneficial and minimise harmful washback, testing policies and types of assessment are the two primary factors that should be the first to intervene.

At student-level, content and method washback and pressure washback are also promoted greatly by the change in testing policies, and teachers' behaviours. So, to promote beneficial washback and minimise harmful ones testing policies, types of assessment and teachers behaviours are the factors that should be given priority. The models discussed above have tried to rationalise that the testing policies, types of assessments, curriculum and resources play concerted role to generate beneficial washback on language teaching and learning. However, Nguyen (ibid.) shows teacher-level washback and student- level washback separately. Though the models seem to be potential in term of washback generation, they are highly ambitious in term of teachers' actual behaviour in the class.

2.5.1.9 Saif's Washback Model

Saif (2006) proposes a model of washback (Figure 2.10) to show how different phenomena such as needs, mean, and consequences work to generate washback on learning. The components of the model systematically represent the major focus areas grouped under three categories of 'needs', 'means', and 'consequences'. The proposed model would allow the inclusion of certain areas of potential impact on the participants of this particular context thereby facilitating the 'washback to the program' (Bailey, 1996). The model illustrates two major lines of connection to be pursued with respect to the test: first, the needs and objectives of the population and the educational context in question, which directly or indirectly affect the type, purpose, and content of the test, its development and implementation; and, second, the potential effects of test use on classroom teaching and learning activities:

Figure 2.10: Saif's washback model (2006)
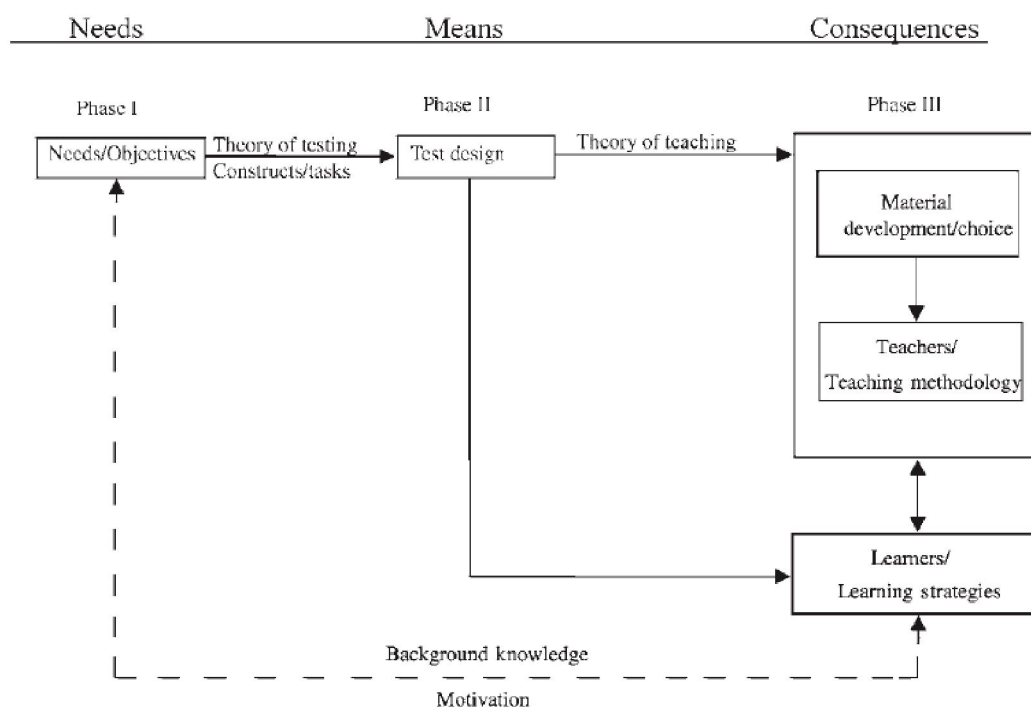
For example, the model suggests that the test be developed with respect to a theoretical framework in conformity with the test objectives so that the same theoretical line of thought can be followed (for example, by teachers and material developers) in all future decisions made with respect to material development and teaching methodology. Moreover, to enhance desirable learning effects, the model suggests that such factors as learners' motivation and background knowledge – previous experience with the target language as well as their topical knowledge – be taken into consideration in the development of the test.

The two-way relationship between the components of the model further allows for what Shohamy (1992) calls the involvement of 'the ones expected to carry out change' (in this case, the teacher) in the test development and/or administration process. Empirical research – with the purpose of examining the possibility of creating washback through the introduction of a new test based on the specific needs of the learners – was then carried out in three different phases each of which corresponded to one of the different levels of the model described above. Saif's model (2006) displays how needs, means, and consequences work in a systematic way under theories of testing and teaching, however, the model does not depict how positive washback can be maximized.

2.5.1.10 Shih's Washback Models

CHIH-MIN SHIH is a researcher in applied linguistics and one of the recognised experts in testing and washback. He proposes two well-known washback models which draw attention of worldwide washback researchers. Shih (2007) proposes a model that describes the roles that both beliefs and contextual factors play in the process of washback (Figure 2.11). The model describes contextual factors as Extrinsic factors which include Socio-economic factors, School and educational factors, Family, friends, and colleagues, and Personal factors. The influence of beliefs appears to be labeled as Personal perceptions of the test under Intrinsic factors. The model below (Figure 2.11) depicts the different roles that both beliefs and contextual factors play:

Figure 2.11: Shih's washback model (2007)
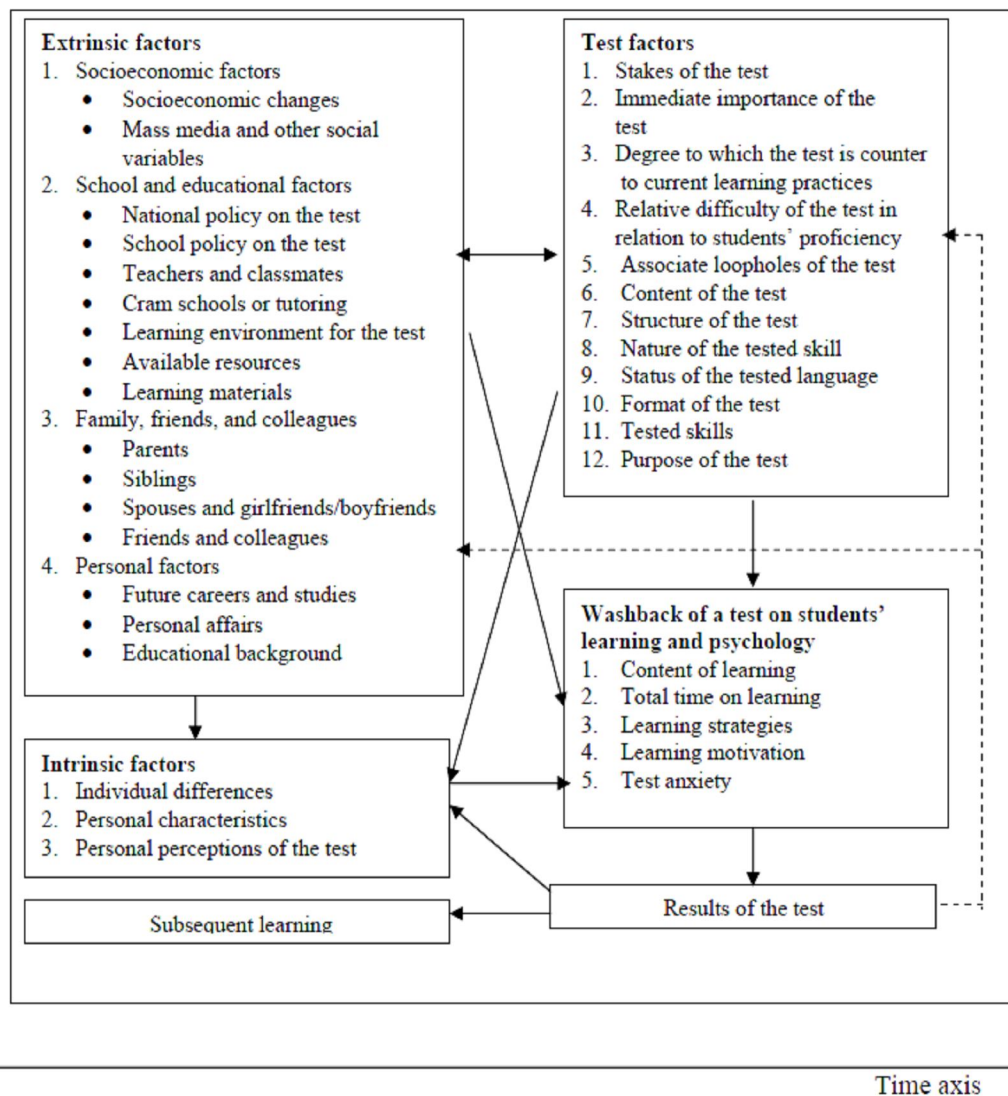
The model includes the *Test factors* as the mediating factors for washback. In the model (Figure 2.11), solid line arrows indicate the impact that has been empirically established and dotted line arrows represent the possible effects which have yet to be investigated. Shih's model (2007) describes not only the direct influence of Extrinsic, Intrinsic, and Test factors on washback, but also their indirect influence on washback. For example, Extrinsic factors can influence washback via Intrinsic factors or Test factors. Test factors can influence washback via Extrinsic factors. One interesting feature of this model is that Shih (2007) includes a time axis to indicate time as a variable, a concept also discussed by Shohamy et al., (1996) who suggests that washback is likely to evolve over time. Shih's model, based on the washback of the General English Proficiency Test (GEPT) on teaching and learning in Taiwan, covers adequate factors. It shows how the factors depend on each other to generate washback.

One concern in Shih's model (2007) is that some items categorised as test factors share similarities such as the content, and test structure, test skills, as well as yet another distinguishing facet that Shih terms "the nature of the tested skills" which are all thought to have some influence on test performance. A more detailed explanation of how these items impact students' learning is also provided. For example, Shih states that test content influences students' learning but does not indicate in what way. It is unclear whether students at the school where the GEPT is a graduation requirement spent more time listening to audio versions of test-preparation materials or not. Another example regarding test impact is that Shih states most students do not prepare for speaking test items because they do not know how to prepare for them. However, he does not clearly reveal the reasons for that.

Shih (2009) has proposed another washback model that builds on that of Bailey (1996). The model is also empirically developed, based on his study of the implementation of the General English Proficiency Test (GEPT) in Taiwan. His data comes from his interviews with participants and in-class observations in Institutions of Higher Learning in that context. The figure below (Figure 2.12) shows that contextual factors, test factors and teacher factors influence the degree of washback on teaching:

66

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

Figure 2.12: Shih's washback model (2009)

Contextual factors (t)

National, social or broader educational factors	**School-level factors**	**Course-level factors**
1. *Socioeconomic climate*	1. *Types of schools (cram school or formal schooling)*	1. *Objectives of the course*
2. *National policy of the test*	2. Resources and support from the school or test designer	2. The size of the class
3. *Competition among schools*	3. The overall program	3. The level/ proficiency of students
4. Test designers	4. *School's policy on the test*	4. The timing of the course (Is the course close to the test date?)
5. *Teaching materials in the market*	5. Management within school	5. <u>Feedback from students</u>
	6. Exam-specific teaching training	6. <u>Students' learning motivation</u>
	7. Resistance from other subject teachers	7. <u>Impact from parents</u>

Washback of tests on teaching (t)
1. *Content of teaching*
2. Teaching method
3. *Teacher-made assessment*
4. *Teacher talk*
5. Time allotment for test preparation
6. *Teacher assigned homework*
7. Nervousness and anxiety
8. Atmosphere of the class

Test factors (t)	**Teacher factors (t)**
1. *The stakes of the test*	1. Teacher training process
2. *The content of the test*	2. Learning experience
3. The extent to which the test is counter to current teaching practices	3. Teachers' abilities in the language they teach
4. *The status of the language tested*	4. Teaching experience
5. The quality of the test	5. Beliefs in effective teaching and test preparation
6. Language skills tested	6. *Perceived quality of the test*
7. The purpose of the test	7. *Teachers' perceived importance of the test to the student*
8. Extra administrative work entailed by the test	8. *The degree of the teachers' commitment to the profession*
	9. Other teaching commitments or various obligations
	10. Teachers' willingness and capability to innovate
	11. Teachers' familiarity with a range of teaching methods
	12. *Teachers' familiarity with the test and the educational context*
	13. *The impact of students' performance on teachers' jobs*
	14. *Teaching philosophy (e.g., objection to test-driven instruction)*

This model focuses only on the washback effect on student learning. In Shih's model (Figure 2.12), the dotted lines denote the impact of one category of factors on another. The symbol (t) acknowledges that washback phenomena may evolve over time, as Shohamy et al. (1996) point out. Factors in italics are either derived from this study or have been reported by other empirical studies, and are substantiated again in my study. Underlined factors have not been corroborated by any empirical data, but it is believed that they are integral to understanding washback. His models largely contribute to academic research though further research is still needed to deepen the understanding of washback.

2.5.1.11 Pan's Washback Model

Pan (2008) proposes a model of washback which seems to be very relevant to EFL education. Her model is generated from the previous analysis of washback studies and the major washback models, and current leading theories such as Alderson and Wall's fifteen washback hypotheses, Bailey's basic model of washback, and Hughes' trichotomy of washback. Her "Micro and Macro-Washback" model is presented in Figure 2.13. This model incorporates ideas from Hughes (1993, as cited in Bailey, 1999) in describing a trichotomy of test effects in terms of "participants", "process", and "product".

Figure 2.13:   Pan's holistic washback model (2008)
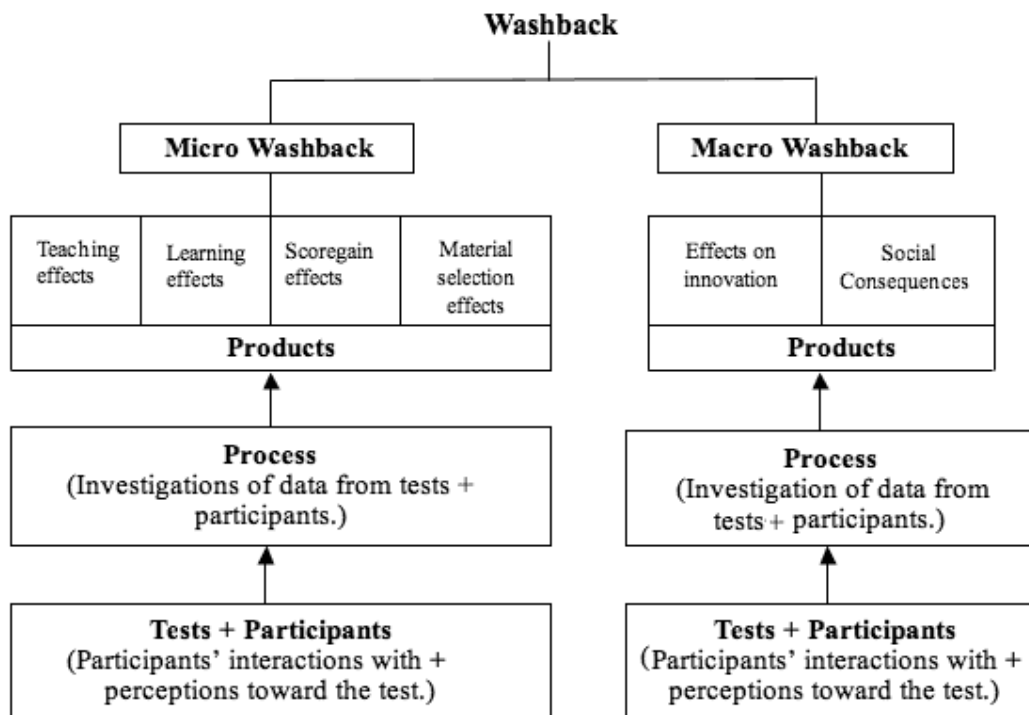
Like other washback researchers, Pan (2008) believes that tests can affect teachers, students, administrators, materials writers, and publishers in terms of their perceptions, activities they engage in, as well as the amount and quality of learning outcomes. Bailey (1996) has combined the fifteen hypotheses from Alderson and Wall (1993) within the trichotomy of the backwash model proposed by Hughes (1993), and created the "basic model of washback" (see figure 2.1). Bailey distinguishes between "washback to the learner" (what and how learners learn and the rate/sequence and degree/depth of learning) and "washback to the program"

(what and how teachers teach and the rate/sequence and degree/depth of teaching) to illustrate the mechanism by which washback works in actual teaching and learning contexts. A common characteristic of these washback models is that they tend to highlight what washback looks like and who is affected, but do little to address the factors that contribute to the phenomenon. In other words, "process" is less understood than "participants" and "products". Besides, the products in these three models/hypotheses refer mainly to teaching and learning washback, not to the aspects of washback that might impact society.

The proposed model in Figure 2.13 aims to strive to represent a holistic balance of both micro-and-macro levels. Washback at the micro level is postulated to consist of teaching, learning, teaching material and score gain effects, while washback at the macro level is postulated to consist of innovation and social dimension features. The different aspects of both levels are viewed as "products", in Hughes's (1993) term. "Tests + Participants", the first item in Figure 2.13, represents participants' (applying Hughes's terms) interactions with and perceptions toward tests, while "process", the second of Hughes's terms and the second item refers to the investigation of data derived from "Tests + Participants" intended to explain those products.

The model investigates how three general phenomena interact on both the macro and micro levels. In addition, this model advocates a well-rounded investigation of washback that focuses not only on a given educational context but also society at large. To gauge micro- and macro washback levels of washback, a triangulation of questionnaires, interviews, observations, pre-and-post tests, and document analysis need to be conducted. This process involves many different stakeholders such as teachers, students, administrators, policy-makers, family members and the general public. The model deserves appreciations as it contributes to further research in applied linguistics.

2.5.1.12 Tsagari's Washback Model

Tsagari (2009) offers a washback model to illustrate the complex ecology of examination washback. In the model (Figure 2.14), washback is represented as an open loop process identifying the number of stakeholders involved in the process and attempting to portray the relationship between them. However, despite it being a

multi-directional relationship among stakeholders, the model, in its visual representation below, is 'simplified' to make it possible to represent it graphically.

Figure 2.14: Tsagari's washback model (2009)

Figure 1: Proposed model of washback
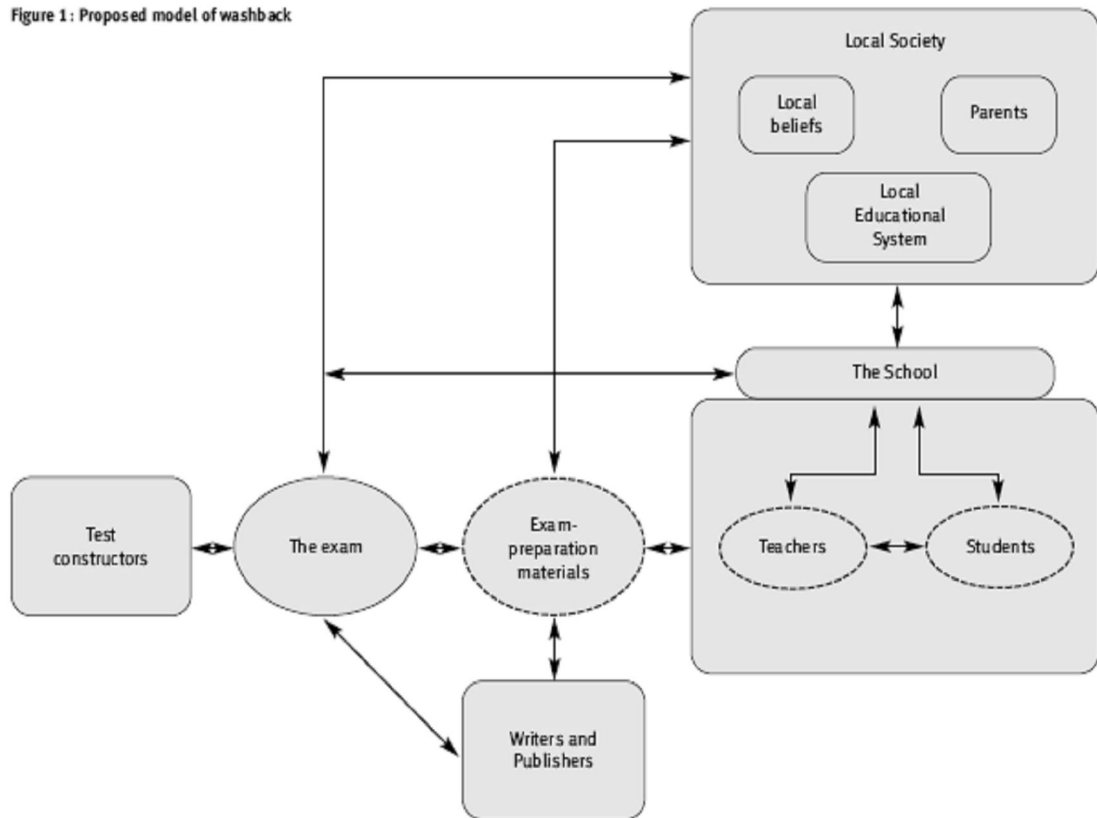
In the above model, the nature of examination washback is circuitous and interactive. The model shows that the examination washback is indirectly engineered on teaching and learning that takes place in the examination-preparation classroom through the understanding of the examination requirements. The model shows that the examination washback is mediated through commercially produced materials that are shaped by the perceptions of the needs of teachers and students by writers and publishers of the materials.

The examination preparation materials mediate between the examination intentions and the examination preparation class. The teacher's role is also crucial in the process as they mediate between material and students. Within this process, washback is also mediated by the school and strengthened by the perceptions and understanding of various other stakeholders operating in the wider local community, such as parents, as well as by the local educational system and beliefs about the

examination and the language tested. Tsagari's model (2009) highlights the process of meditation of washback through the use of materials. If he would in corporate the role of teaching methodology, the model could be more acceptable.

2.5.1.13 Mizutani's Washback Model

Mizutani (2009) proposes a washback model demonstrating that certain types of washback effects are mediated by certain types of contextual factors and beliefs. In her proposed model, beliefs are illustrated as having a direct influence on the way the nature of washback is interpreted:

Figure 2.15: Mizutani's washback model (2009)
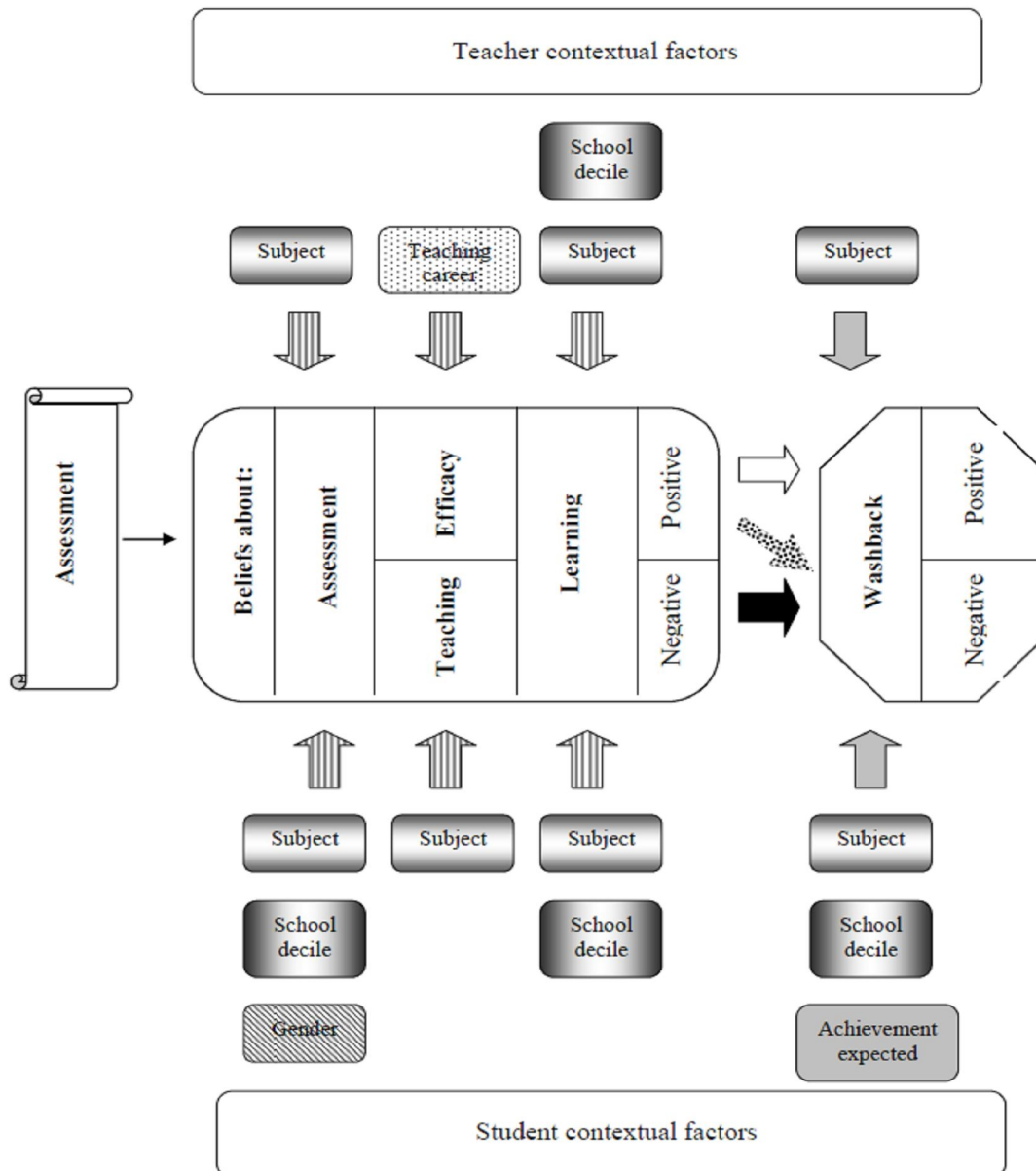
In the figure above, (Figure 2.15), the white block arrow shows that beliefs that are positive are likely to bring about positive washback while the black block arrow indicates that beliefs that are negative are likely to cause negative washback. Furthermore, beliefs that are positive can mitigate negative washback, which is indicated by a dot-shaded arrow in the model. Although these patterns were often found in common among teachers and students, more opportunities to promote positive washback and to cause negative washback existed for students.

The two grey block arrows signify the direct influence of contextual factors of teachers and students on washback. For teachers and students direct effects are more subject related, indicating the distinction between verbal or numeric subjects is likely to determine the nature of washback. School decile and achievement expected are also shown to have a direct influence on washback for students. Whether students are from lower or higher decile schools and whether they consider themselves as lower or higher achieving are further factors which are likely to determine the nature of the washback.

Mizutani (2009) suggests that certain contextual factors are likely to influence particular types of beliefs. The striped block arrows illustrate patterns of these influences. For teachers and students, whether their subject is verbal or numeric may influence a certain type of belief about assessment. For students, whether they are from lower or higher decile schools or they are male or female may also influence certain types of belief about assessment. For teachers and students, whether their subject is verbal or numeric and whether they are from lower or higher decile schools may also influence beliefs about learning. For students alone, their beliefs about teaching are likely to be influenced by the subject. Teachers' views about their own efficacy are likely to be influenced by the length of their teaching career. The nature of washback depends on whether these types of beliefs are positive or negative. Thus, certain contextual factors arguably influence washback indirectly via beliefs.

Mizutani's (2009) washback model shows that washback and beliefs are more context-dependent for students than for teachers, while demonstrating similarities between teachers and students in the extent to which contextual factors and beliefs play a role in the process of washback. She claims that it is possible to promote intended positive washback where teacher's and student's beliefs are

aligned with the intentions of the Ministry of Education, and where contexts are supportive. She further confirms that the links established between teachers' and students' beliefs, their contextual factors, and washback in the proposed model are arguably useful to increase understanding of the mechanism of washback of an assessment on teaching and learning. She believes that by clarifying the link between assessment and desired outcomes, the model can potentially help promote intended positive washback while minimising undesirable negative washback. The present researcher finds that the model proposed by Mizutani (2009) seems to be potential to large extent to generate beneficial washback on teaching and learning, however, the proposed model would be more prospective if she could explain how external pressure and test contents contribute to the generation of washback.

This section looks at the functions and mechanisms by which washback works in relation to other educational theories and practices. Washback is a complex phenomenon. Similarly, the models which have been proposed during last 20 years are not clearly defined because of its variability. The washback models discussed above have been designed in different educational context. The researchers have proposed washback models on the basis of their own contexts. Thus, the current research set out to develop a washback model which could describe the way washback was mediated particularly by beliefs held by both teachers and students and their contextual factors. Future washback research would probably benefit from incorporating theories of test impact available in both fields.

2.6 Areas Affected by Washback

The view of testing is derived from the realisation of test power and its manifestations with regard to high-stakes decisions based on test results for individuals, educational systems and society as a whole. Many research studies reveal that a test affects participants, processes, and products in teaching and learning. Students, teachers, administrators, material developers and textbook writers may be included under the term 'participants'. Their perceptions and attitudes towards their work are likely to be affected by a test. Process refers to any action taken by the participants, which may contribute to the process of learning. Material development, syllabus design, use of syllabus and curriculum, applying teaching methodology, and the use of learning and/or test-taking strategies are

included under processes. Product means what is learned (facts, skills, etc.) and the quality of the learning (e.g. fluency, competence, etc.). Tests have an impact on the learning outcomes as well.

As mentioned, washback affects various aspects of teaching and learning, such as syllabus and curriculum, stakeholders, materials, teaching methods, testing and mediating factors, learning outcomes, feelings, attitudes, and learning, etc. Tests have impact on the lives of test takers, classrooms, school systems and even whole societies (Hamp-Lyons, 1998). Wall & Alderson (1993) put forward the 15 hypotheses, highlighting more specifically some of the ways in which a test might affect teaching and learning. The five of the hypotheses relate to washback to the learners, six relate to washback to the programme, and four relate to syllabus, curriculum, and teaching contents. Their hypotheses are:

Hypothesis	Relates to:
1 A test will influence teaching. --	Teachers
2 A test will influence learning. --	Learners
3 A test will influence what teachers teach; and ----------------------------	Teachers
4 A test will influence how teachers teach; and therefore by extension from (2) above: --	Teachers
5 A test will influence what learners learn; and ----------------------------	Learners
6 A test will influence how learners learn. ----------------------------------	Learners
7 A test will influence the rate and sequence of teaching; and -------------	Teachers
8 A test will influence the rate and sequence of learning. ------------------	Learners
9 A test will influence the degree and depth of teaching; and --------------	Teachers
10 A test will influence the degree and depth of learning. ------------------	Learners
11 A test will influence attitudes to the content, method, etc. of teaching and learning. --	Teachers & learners
12 Tests that have important consequences will have washback; and conversely.	High stakes tests
13 Tests that do not have important consequences will have no washback.	Low stakes tests
14 Tests will have washback on all learners and teachers. -------------------	Teachers & learners
15 Tests will have washback effects for some learners and some teachers, but not for others.---	Teachers & learners

The Washback Hypothesis seems to assume that teachers and learners do things they would not necessarily otherwise do because of the test. Additionally, in

order to study the washback effect, it is necessary to look at the people that participate in the educational process, to the actual classroom events and activities, and to the outcomes of these processes. Based on the various types of research throughout the world, washback hypotheses may be summarised as:

1. Tests can affect curriculum and learning,
2. Tests can provide feedback on learning,
3. Tests can help implement content and performance standards,
4. Tests can influence the methodology that teachers use,
5. Tests can motivate teachers and students,
6. Tests can orient students as to what is important to learn,
7. Tests can help orient needed teacher training,
8. Tests can help implement articulation,
9. Tests can help implement educational reform.

A curriculum is a vital part of the EFL classes, and washback has deep relation with the syllabus and curriculum. Test contents can have a very direct washback effect upon teaching curricula. It provides a focus for the class and sets goals for the students throughout their study. A curriculum also gives the student a guide and idea to what they will learn, and how they have progressed when the course is over. The test leads to the narrowing of contents in the curriculum. Tests can affect curriculum and learning (Alderson & Wall, 1993). Shohamy et al. define curriculum alignment as "the curriculum is modified according to test results" (1996, P.6). The findings from the studies about washback onto the curriculum indicate that it operates in different ways in different situations, and that in some situations in may not operate at all.

Learners follow a 'hidden' syllabus, that is, the contents driven by the contents of examination. Alderson and Wall (1993) conclude from their Sri Lanka study that 'the examination has had a demonstrable effect on the content of language lessons' (p, 126-127). This effect is that of the narrowing of the curriculum to those areas most likely to be tested. This finding is similar to that of Lam (1994) who has reported an emphasis in teaching on those parts of the exam carrying the most marks. The findings of Read and Hayes (2003) are quite detailed and show variations in washback on the curriculum depending on the course observed. The studies discuss the effects of washback on various aspects of the classroom, which

can be categorized as follows: curriculum, materials, teaching methods, feelings and attitudes, learning. This section reviews the findings for each of these areas in turn.

2.6.1 Washback on Syllabuses and Curriculums

Many researchers (e.g. Bailey, 1996, 1999; Wall & Alderson, 1993, Wang, 2010; Hsu, 2009) of high-stakes tests attest that tests are responsible for narrowing the school curriculum by directing teachers to focus only on those subjects and skills that are included in the examinations. As a consequence, such tests are said to "dominate and distort the whole curriculum" (Vernon, 1956: 166; see also Kirkland, 1971; Shepard, 1991). A test was considered to have beneficial washback, when preparation for it did not dominate teaching and learning activities narrowing the curriculum. When a test reflected the aims and the syllabus of the course, it was likely to have beneficial washback, but when the test was at variance with the aims and the syllabus, it was likely to have harmful washback.

Wall & Alderson (1993) put forward the 15 hypotheses, highlighting more specifically some of the ways in which a test might affect teaching and learning. The following are the hypotheses that relate to syllabus, curriculum, and teaching contents:

 (3) A test will influence what teachers teach; and
 (5) A test will influence what learners learn; and
(7) A test will influence the rate and sequence of teaching (P); and
(11) A test will influence attitudes to the content, method, etc. of teaching and learning (ibid).

Examination should reflect the syllabus and curriculum, and since not everything in a curriculum can be tested in an examination, the areas that are assessed should be ones that are considered important. It is also important that, same items and contents should not be tested again and again. Insofar as possible, modes of testing (e.g., written, practical, oral) should be diverse to reflect the goals of curricula. The format and contents of the public examination should be reorganized every year. The use of commercially produced clone tests materials in the class should be discouraged. Teaching to the test universally occurs in either the practice of frontloading or backloading. If a high match exists between the curriculum and

the test, teaching to the test is inevitable and desired. Otherwise, the data produced by the test is not useful in improving teaching and learning. In this case, using tests as the source to develop curriculum runs the risk of accepting and defining learning only in terms of what is tested in the test.

2.6.1.1 Alignment of Curriculums with Public Examinations

A curriculum provides a focus for the class and sets goals for the students throughout their study. A curriculum also gives the student a guide and idea to what they will learn, and how they have progressed when the course is over. Examinations or high-stakes tests exert a considerable impact on what, and how, teaching and learning are conducted in the classroom. Alderson and Wall (1993) elaborate, saying that "for teachers, the fear of poor results, and the associated guilt, shame, or embarrassment, might lead to the desire for their pupil to achieve high scores in whatever way seems possible. They point out this might lead to 'teaching to the test', with an undesirable narrowing of the curriculum" (ibid. p.118).

Alignment of the curriculum refers to the match between the content and format of the curriculum and the content and format of the test. Curriculum alignment is a process to improve the match between the formal instruction that often occurs in the classroom and the instrument that is used to measure the instruction outcomes. It is now proven fact that washback has a deep relation with the syllabus and curriculum. Test contents can have a very direct washback effect upon teaching curricula. Tests can affect curriculum and learning (Alderson & Wall, 1993). Shohamy et al. define curriculum alignment as "the curriculum is modified according to test results" (1996, p.6).

A curriculum is a vital part of TEFL classes. It provides a focus for the class and sets goals for the students throughout their study. A curriculum also gives the student a guide and idea to what they will learn, and how they have progressed when the course is over. *Curriculum alignment* focuses on the connection between the testing and teaching syllabus (Andrews, 1994; Madaus, 1988; Shepard, 1993). Systemic validity implies the integration of tests into the educational system and the need to demonstrate that the introduction of a new test can improve learning (Cheng, 1997). Frederiksen & Collins (1989: 27) state that 'A systematically valid test is one that induces in the education system curricular and instructional changes that foster

the development of the cognitive skills that the test is designed to measure'. Pierce (1992) states "the washback effect, sometimes referred to as the systemic validity of a test (p.687). The test leads to the narrowing of contents in the curriculum:

Figure 2.16: Washback on syllabus and curriculum by Saville& Hawkey (2004)
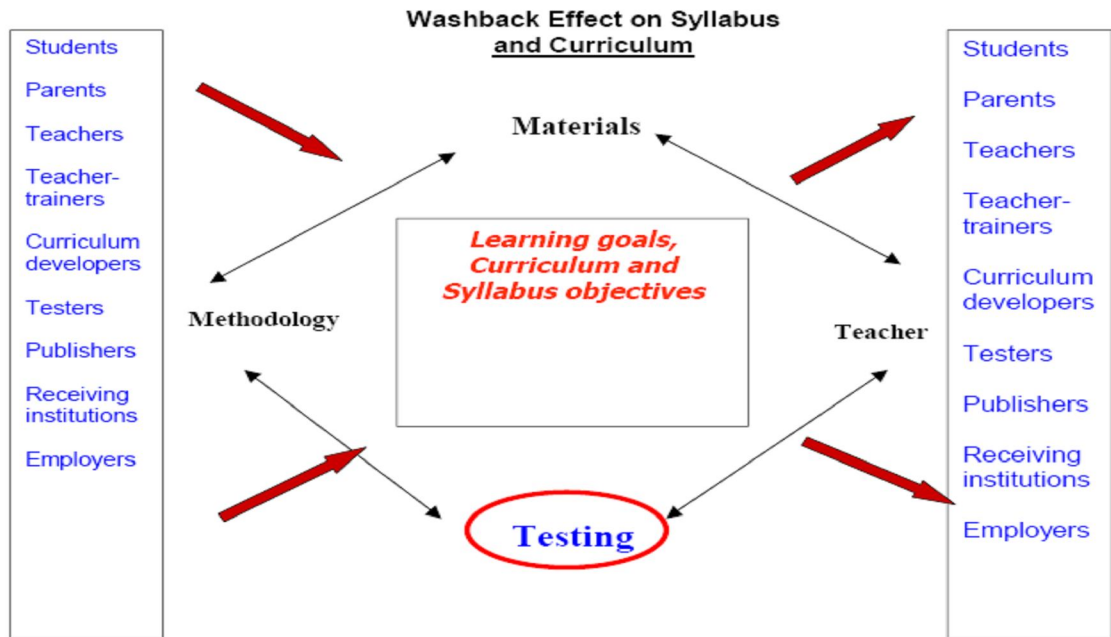
Curriculum alignment is commonly regarded as a process to improve instruction and tests. The process of curriculum alignment is usually established by two ways, frontloading and backloading.

2.6.1.2 Curriculum Alignment by Frontloading

Frontloading alignment is commonly practiced in education. It is assumed that frontloading can prevent teaching to the test, which may lead to an extremely narrow and rigid view of the actual goals and objectives of any curriculum. In the process of frontloading alignment, the curriculum is developed first and the test is designed to measure or assess whether students have learned what the curriculum includes. In this scenario, the test always follows and does not lead the curriculum (Lindvall and Nitko, 1975). Given an inappropriate test, narrowing of curriculum impedes teaching and learning (Smith, 1991).

2.6.1.3 Curriculum Alignment by Backloading

Opposite to frontloading, backloading refers to working from the test back to the curriculum, in terms that the curriculum to be taught is derived from the test to be given. (Table 2.2) It is assumed that backloading alignment can produce quick results in improved test scores (Niedermeyer and Yelon, 1981). However, issues of teaching to the test remain the most troublesome problem in the whole backloading alignment process. One issue is whether anything on the instrument that ought not to be taught is tested. The other issue, a local educator often asks, is whether anything that a student should know is not tested or assessed. The table below (Table 2.2) illustrates the process of Frontloading vs. backloading curriculum alignment (Steffy, 2001):

Table 2.2: Frontloading vs. backloading process of curriculum alignment

Frontloading	Design	Delivery
	Write the curriculum first and then develop a test to assess it.	Teach the curriculum first and develop a test to assess it.
Backloading	Obtain publicly released test items and create a curriculum based upon them.	Obtain publicly released test and create parallel classroom structures in which content/ is embedded.

It is common to claim the existence of washback (the impact of a test on teaching) and to declare that tests can be powerful determiners, both positively and negatively, of what happens in classrooms. One of its key characteristics is the careful observation of teacher behavior. Swain (1985) says "It has frequently been noted that teachers will teach to a test: that is, if they know the content of a test and/or the format of a test, they will teach their students accordingly" (p. 43). It is generally accepted that public examinations influence the attitudes, behavior, and motivation of teachers, learners and parents (Pearson, 1988).

Tests are often perceived as exerting a conservative force which impedes progress. Andrews and Fullilove point out, "Not only have many tests failed to change, but they have continued to exert a powerful negative washback effect on teaching" (Andrews and Fullilove, 1994, p. 57). Heyneman (1987) has commented that teachers teach to an examination. Alderson and Wall (1993) concluded from their Sri Lanka study that 'the examination has had a demonstrable effect on the content of language lessons' (p, 126-27). Lam (1994) finds that more curriculum

time is given to exam classes, though Shohamy et al. (1996) suggest that this is true only in the case of exams viewed as high stakes. Alderson and Hamp Lyons (1996) note in their study that while extra time is given to TOEFL classes in some institutions this is not the case in others.

The findings of Read and Hayes (2003) are quite detailed and show variations in washback on the curriculum depending on the course observed. Pierce (1992, p. 687) specifies classroom pedagogy, curriculum development, and educational policy as the areas where washback has an effect. On the other hand, Alderson and Hamp-Lyons (1996) take a view of washback which concentrated more on the effect of the test on teaching. They has referred to washback as "… the influence that writers of language testing, syllabus design and language teaching believe a test will have on the teaching that precedes it" (ibid: p. 280). Washback can be seen to have operation on teaching content, preparation for tests like training in test taking strategies, doing exercises of past papers, teaching methods, assessing students, and changing curriculum and materials used. Empirical findings are summarised by flowchart in Figure 2.17 below:

Figure 2.17: Washback effect and the possible factors (Pan, 2009)
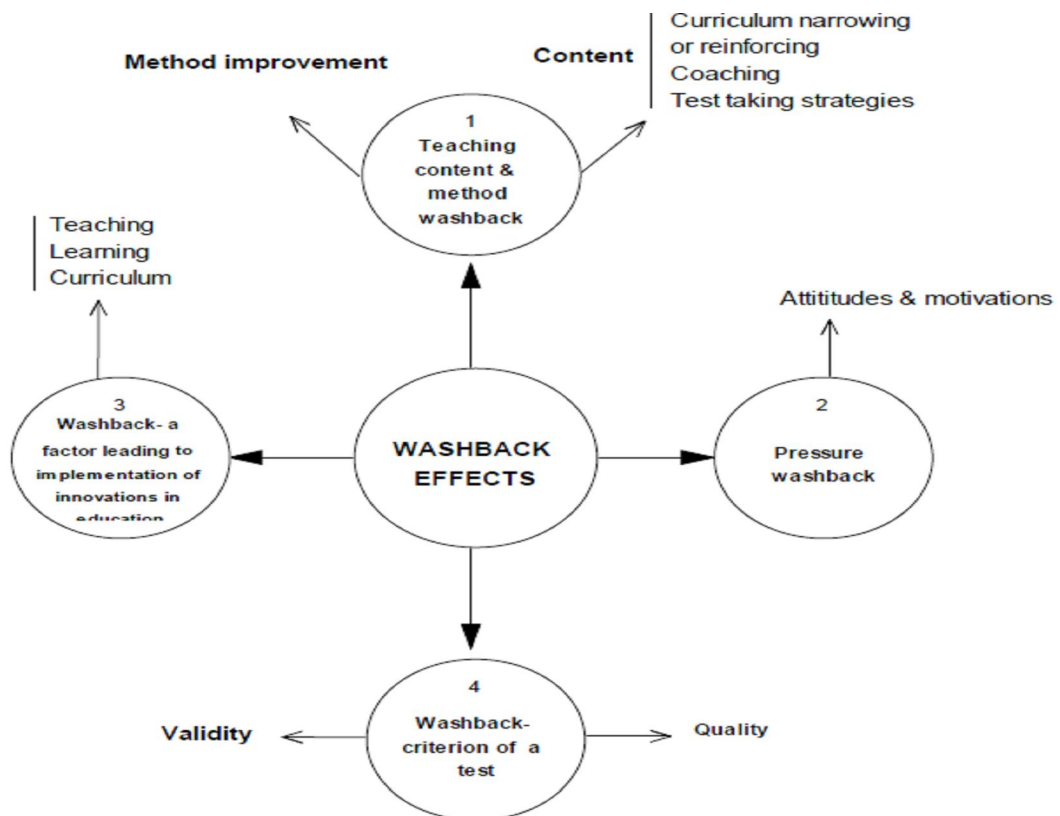
Higher Secondary Learners in Bangladesh follow a 'hidden' syllabus (e.g. past questions, guidebooks), that is, the contents driven by the contents of EFL examination. Cohen (1994) describes washback in terms of" how assessment instruments affect educational practices and beliefs" (p. 41). Bailey's (1999) extensive summary of the current research on language testing washback highlights various perspectives and provides deeper insight into the complexity of this phenomenon. But today, a new perspective (and a new education buzz phrase) is emerging. It's called curriculum alignment, and it means teaching knowledge and skills that are assessed by tests designed largely around academic standards set by the state. In other words, teaching to the test. Alderson and Hamp-Lyons summarise some typical concerns regarding negative washback to the curriculum (1996, p. 28):

1. Narrowing of the curriculum (Madaus, 1988; Cooley, 1991)
2. Lost instructional time (Smith et al., 1989)
3. Reduced emphasis on skills that require complex thinking or problem-solving (Fredericksen, 1984; Darling-Hammond and Wise, 1985)
4. Test score 'pollution', or increases in test scores without an accompanying rise in ability in the construct being tested (Haladyna, Nolan and Haas, 1991)

Spolsky (1994, p. 55) define backwash as a "term better applied only to accidental side-effects of examinations, and not to those effects intended when the first purpose of the examination is control of the curriculum", and spoke of the "…inevitable outcome in narrowing the educational process…" (ibid.). He uses vocabulary tests to illustrate what he calls the 'crux of the backwash problem'. While vocabulary tests may be a quick measure of language proficiency, once they are established as the only form of assessment, the backwash to instruction resulted in the tests becoming a measure of vocabulary learning rather than language proficiency. Negative washback occurs when the test items are based on an outdated view of language, which bears little relationship to the teaching curriculum (ibid.).

Similarly, Wall and Alderson (1993) reason that if the aims, activities, or marking criteria of the textbook and the exam contain no conflicts and the teachers

accept and work towards these goals, then this is a form of positive washback. Negative washback would be evidenced in the exam having a distorting or restraining influence on what is being taught and how. Alderson and Banerjee (2001) acknowledge that tests have the potential to be 'levers for change' in education if one accepts the argument that if bad tests have a negative impact then it should be possible for a good test to have good washback.

2.6.1.4 Teaching to the Test

Teaching to the test--the very words has always been heresy to educators. 'Teaching to the test' puts too much emphasis on standardized tests that are poorly constructed and largely irrelevant, the theory goes; it stifles creativity and encourages cheating. Vallette (1994) suggests that washback is particularly strong in situations where the students' performance on a test determines future career options. In such case, teachers often feel obliged to teach to the test, especially if their effectiveness as a teacher is evaluated by how well their students perform.

The assumption that frontloading alignment prevents teaching to the test is often not the case, in terms that teaching to the test still occurs under the practice of frontloading. If the curriculum and the test correspond to each other, teaching to the test is inevitable and desired. The extent to which a test is useful to a given curriculum is the extent to which the test indeed measures the curriculum in the first place. In the alignment by frontloading, examining the test itself is one way to assess the test quality, in terms of determining whether anything on the instrument that ought not to be taught is tested or that ought to be taught is not tested. A backloaded curriculum assumes "null curriculum"; that is, the content not tested or assessed in the test is not included in the curriculum. The act of "null curriculum" or "non-selection" is valued laden. The values not selected by the test makers represent an unknown element that may be at odds with local values.

2.6.2 Washback on Teaching Methodology

By teaching methods the present researcher refers to teaching approaches or techniques. The findings on this area are once again not homogeneous. While Alderson and Wall (1993, p. 127) says that their Sri Lanka study showed the exam

'had virtually no impact on the way that teachers teach'. Andrews et al. (2002) point out that the revised exam led to teachers' use of explanation of techniques for engaging in certain exam tasks.

Cheng (1997) mentions that teaching methods may remain unchanged even though activities change as a result of the revision of an exam; in this case reading aloud was replaced by role plays but both were taught through drilling (p, 52). The high-stakes EFL examination leads teachers to teach through simulating the examination tasks or through carrying out other activities that directly aim at developing exam skills or strategies (e.g., brainstorming, working in pairs or in groups, jigsaw activities, simulating authentic situations, engaging in debates, discussions, speeches, etc.). Watanabe's findings for this area are once again different. He reports that the teachers in his study 'claimed that they deliberately avoided referring to test taking techniques, since they believed that actual English skills would lead to students' passing the exam' (2000, p. 45).

Some of the studies indicate that the methods used to teach towards exams vary from teacher to teacher. Alderson and Hamp Lyons (1996), and by Watanabe (1996) find large differences in the way teachers teach towards the same exam or exam skill, with some adopting much more overt 'teaching to the test', 'textbook slave' approaches, while others adopted more creative and independent approaches (p, 292).The researchers in both these studies stress that the variable may be not so much the exam or exam skill as the teacher him=herself. They go on to discuss various teacher-related factors that may affect why and how a teacher works towards an exam. Teacher attitude towards an exam would seem to play an important role in determining the choice of methods used to teach exam classes. There has been a perception that washback affects teaching content and teaching methods. It seems to be true in some circumstances but not others, suggesting that whether the exam affects methods or not may also depend on factors other than the exam itself, such as the individual teacher. Other findings on teaching methods relate to interaction in the classroom.

Alderson and Hamp Lyons (1996) note in their investigation of TOEFL teaching that the exam classes spend much less time on pair work, that teachers talk more and students less, that there is less turn taking, and the turns are somewhat longer. Watanabe (2004) notes that 'students rarely asked questions even during

exam preparation lessons'. Cheng (1998) points out that while teachers talk less to the whole class as a result of the revised exam, the teacher talking to the whole class remains the dominant mode of interaction.

It is seen that examination oriented materials are heavily used in classrooms particularly when the examination approaches. However, it is not clear from the studies that it is the exam that generates less interaction in exam classes, or whether this is due to teachers believing, for whatever reason, that this is the way exams should be prepared for. The type and amount of washback on teaching methods appears to vary from context to context and teacher to teacher. It varies from no reported washback to considerable washback. The variable in these differences appears to be not so much the examination itself as the teacher.

2.6.3 Washback on Teacher Factors

Teacher perception, teacher attitudes and teacher beliefs are often mentioned in the washback studies as powerful factors. Among the factors that can mediate the washback effect is the teacher (Wall, 1996) and her/his perceptions about the examination, its nature, purposes, relevance in the context, etc. What have been noted in the results are the behaviors of teachers in response to examination changes. However, as Shavelson and Stern (1981) argue, examining only teacher behavior is incomplete. There is a need to examine the link between teacher intentions or beliefs and how this translates into action (Tsui, 2003; Woods, 1996). By doing so, predictable variations in teachers' behaviour that result from differences in goals, judgments and decisions can be better accounted for. According to Shulman (1986, 1987) research that links teachers' intentions to their behaviour provide a sound basis for educating teachers and implementing educational innovations.

It is argued that the dictates of high-stakes tests reduce the professional knowledge and status of teachers and exercise a great deal of pressure on them to improve test scores which eventually makes teachers experience negative feelings of shame, embarrassment, guilt, anxiety and anger. Green (2006, 2007) starts to examine this facet of washback. Johnson (1992), Sato and Kleinsasser (1999), Tan (2008), Turner (2006, 2008) and Wang (2008) have showed that teacher factors influence teaching practices in the classroom. Teacher beliefs are consistent with

their prior experience and instructional approaches. There is, therefore, an increasing realisation in the field of assessment that the "teacher factor" is fundamental to the kind of washback effect that takes place in the classroom.

Wall and Alderson (1993) comment the examination has considerable impact on the content of English lessons and on the way teachers designed their classroom tests (some of this was positive and some negative), but it has little to no impact on the methodology they used in the classroom or on the way they marked their pupils' test performance. Among many important results of the Sri Lankan impact study, Wall and Alderson make the following summary statements about the impact of the new Sri Lankan texts and tests on the teachers (ibid., p. 67):

1. A considerable number of teachers do not understand the philosophy/approach of the textbook. Many have not received adequate training and do not find that the *Teacher's Guides* on their own give enough guidance.

2. Many teachers are unable, or feel unable, to implement the recommended methodology. They either lack the skills or feel factors in their teaching situation prevent them from teaching the way they understood they should.

3. Many teachers are not aware of the nature of the exam- what is really being tested. They may never have received the official exam support documents or attended training sessions that would explain the skills students need to succeed at various exam tasks.

4. All teachers seem willing to go along with the demands of the exam (if only they knew what they were).

5. Many teachers are unable, or feel unable, to prepare their students for everything that might appear on the exam.

Watanabe (2004a) finds that the presence of grammar translation questions on a particular university entrance exam did not influence these two teachers in the same way. He has identified three possible factors that might promote or inhibit washback to the teachers: (1) the teachers' educational background and/or experiences; (2) differences in teachers' beliefs about effective teaching methods; and (3) the timing of the researcher's observations. (Teacher A was observed when the exams for which the students were preparing were six months away, while Teacher B was teaching exam-preparation classes just a month or so before the

entrance examinations would occur.) Thus Watanabe concludes that "teacher factors may outweigh the influence of an examination" (ibid., p. 331) in terms of how exam preparation courses are actually taught.

Tests can aid learning and teaching both if aimed to assess the required skills. Many researches have been carried out on washback explicating that it can be either beneficial or harmful depending upon the contents and techniques (Alderson & Wall 1993; Bailey 1996, p. 257; Cheng & Falvey 2000). For example, if skills not required for every day communication are assessed, the test could leave harmful effect on teaching and learning, such as mechanical test of writing skills by giving multiple-choice questions on grammar. A great number of washback studies (e.g. Alderson & Hamp-Lyons, Cheng, 2004; Cheng, 2005; Ferman, 2004; Hawkey, 2006; Lam, 1994; Qi, 2005; Saif, 2006; Wall & Horak, 2006; Watanabe, 1996; Watanabe, 2004; Shih, 2007; Pan, 2009) focus on what takes place in the language classroom. Many researchers (e.g. Cheng 2004, Wall & Alderson, 1993; Turner 2007; Qi, 2004, 2005) find that content changes because of the test, but the way teachers instruct does not vary to any great degree. The changes were "superficial" (Cheng, 2005, p. 235), not substantial.

A majority of teachers tended to "teach to the test." For example, Green (2006, 2007) and Hayes & Read (2003,2004) find more test-related activities (e.g. offering test-taking tips, doing question analysis) in the IELTS preparation classes than in the EAP (English for academic purpose) classes. In addition, teachers' beliefs and attitudes regarding the immediate goals of teaching and their own limited ability to use the language effectively contribute to their being unable to effect the positive changes (a shift in English language teaching to a more communicative orientation) the test developers intended to create (Qi, 2005). Cheng (2004) asserts that inadequate training and teachers' professional backgrounds lead to unchanged methodologies because they don't know how to change, not that they do not want to change.

A good number of researchers (e.g. Alderson & Hamp-Lyons, 1996; Watanabe, 1996; Wang; Shih, 2010), however, find that tests affected both how and what teachers taught but not all teachers reacted the same way to the same test. In many instances, teachers reported a greater sense of pressure from the tests (Watanabe, 2004b; Burrows, 1998; 2004). Shohamy (1993); and Shohamy et al.

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.

86
<Pages 288-786>

(1996) also have discovered significant differences between experienced and novice teachers. The former tends to teach to the test and uses only material to be included in the test, while the latter uses different activities to teach oral language. Lam (1994) has reported that more experienced teachers tend to be significantly more "examination-oriented" (p. 91) than their younger colleagues. The new teachers are found more sincere language teachers than the experienced or older ones. The more the teachers get experienced, the more the teachers teach to the test. The experienced teachers are relatively misguided by the examination, and thus, create very negative washback on their teaching.

The findings of the previous studies on teaching show that washback are contradictory in terms of what (content) and how (methodology) teachers teach. This may be attributed to Hawkey's claim (2006) that "the distinction between course content and methodology is not always clear cut" (p. 106). Nevertheless, researchers (Burrows, 2004; Cheng, 1997; Wall & Alderson, 1993; Watanabe, 1996; Watanabe, 2004b) seem to have reached a consensus on the concept that tests influence what happens in the classroom in terms of teaching activities and content, and that teachers' beliefs, and educational backgrounds play an important role in deciding how they instruct the students in the class.

2.6.4 Washback on Language Learning

There is a general understanding that washback is a complex phenomenon. Many researchers call for empirical studies to explore the concept further. It is encouraging to note that more and more researchers have expanded to look at issues of context in order to capture the complexity of the washback phenomenon (Cheng, 2001; Cheng, 2004; Davison, 2008; Qi, 2005; Shohamy, 1993; Hamp-Lyons and Tavares, 2008; Turner 2008, 2009; Urmston & Fang, 2008; Wall, 1999; Watanabe, 1996, 2004b). It is obvious that the washback phenomenon has been examined much more seriously, both theoretically and empirically. In comparison to washback studies in other areas, fewer researches have been conducted to investigate the washback effects on students' learning processes. Watanabe (2004) states, "relatively well explored is the area of washback to the program, while less emphasis has been given to learners" (p. 22). Those studies that have been focused on learning washback received varied and sometimes contradictory findings.

Shohamy et al. (1996) contend that an important test promotes learning, while Cheng (1998) shares a similar finding by saying that tests motivated students to learn but that their learning strategies did not change significantly from one test to another. The recent study, Stoneman (2006) investigates how students prepare for an exit examination in Hong Kong. The results show that students are motivated more and spent more time in preparing for higher-status examination (IELTS) than the lower-status test (GSLAP), but preparation methods are much the same. Wall and Alderson (1993) suggest that future washback studies should investigate how tests affect students' motivation and performance. Wall (2000) contends, "What is missing … are analyses of test results which indicate whether students have learnt more or learned better because they have studied for a particular test" (p. 502). To better understand what washback occurs within the classroom, researchers need to investigate changes in students' motivations, learning styles, and learning strategies (Stoneman, 2006).

2.6.5 Washback on Test Takers

The learners are the key participants whose lives are most directly influenced by language testing washback. The washback influences the test takers directly by affecting language learning (or non-learning), while the influences on other stakeholders will affect efforts to promote language learning. The test-takers themselves can be affected by: the experience of taking and, in some cases, of preparing for the test; the feedback they receive about their performance on the test; and; the decisions that may be made about them on the basis of the test. Of the 15-washback hypotheses of Alderson and Wall's (1993, pp. 120-121), five are directly addressed learner washback. Bailey (1996) suggests students face with an important test they may participate in (but are not limited to) the following processes:

1. Practicing items similar in format to those on the test.
2. Studying vocabulary and grammar rules.
3. Participating in interactive language practice (e.g., target language conversations).
4. Reading widely in the target language.
5. Listening to non-interactive language (radio, television, practice tapes, etc.).

6. Applying test-taking strategies.

7. Enrolling in test-preparation courses.

8. Requesting guidance in their studying and feedback on their performance.

9. Requesting or demanding unscheduled tutorials or test-preparation classes (in addition to or in lieu of other language classes).

10. Skipping language classes to study for the test. (pp. 264-265)

Learner washback has also important financial implications for pupils and their families, in terms of their access to educational opportunities. For example, Wall and Alderson examined a context in which a new national test was implemented, this time the O-level exams administered at the end of the 1lth year of education in Sri Lanka. These authors report, "a student's O-level grades, particularly in English, are among the most important in his or her academic career" (1993, p. 42). Washback may affect learners' actions and/or their perceptions, and such perceptions may have wide ranging consequences. Sturman used a combination of qualitative and quantitative data to investigate students' reactions to registration and placement procedures at two English-language schools in Japan. The placement procedures included a written test and an interview. He found that the students' perceptions of the accuracy of the placement.

2.6.6 Washback on Materials

The term 'material' is used here to refer to the prescribed textbooks, guidebooks and past question papers. Examination-related textbooks and other materials can vary in their type of contents. Very often, tests promote a boom of test related materials, and thus, influence what teachers teach in the classroom, but tests may also encourage teachers to use additional materials from a variety of sources. They range on the one hand from materials that are highly exam technique oriented, and make heavy use of parallel exam forms, to those on the other hand that attempt to develop relevant language skills and language. A teacher's choice of materials relies on a number of factors such as the purpose of the test and the availability of ready-made materials. Generally, the studies refer particularly to those materials at the 'highly exam oriented' end of the spectrum.

A large number of studies discuss washback on materials in terms of materials production, the use of materials, student and teachers' views of exam materials, and the content of materials. Most teachers know from their own experience of the rows of exam-related materials available on the shelves of bookshops and staff rooms, and of the new editions of course books and other exam materials that are issued when exams are revised. They find that in relation to the EFL exam 'ample new material has been published and marketed since the announcement of the test changes became public.

Teachers' use of materials seems to vary to large extent. Lam (1994) speaks of teachers as 'textbook slaves' and 'exam slaves' (p.91). He finds that large numbers of teachers rely heavily on the textbook in exam classes, and more heavily on past papers. Lam (1994) also reports that teachers do this, as they believe that the best way to prepare students for exams is by doing past papers. Andrews, et al. (2002) speak of the large role played by published materials in the Hong Kong classroom, citing a previous study by Andrews (1995) in which the teacher respondents were found to spend an estimated two-thirds of class time working on exam-related published materials. Cheng (1997) suggests that a reason for this may be that the exam textbooks in Hong Kong not only provide information and activities but also suggested methods for teaching and suggested time allocations.

The researchers such as Fullilove (1992), Xiaoju (1992), Wall and Alderson (1993), Lam (1994) and Cheng (1997) suggest that test requirements may promote test-related materials, and that these materials affect what teachers instruct because they tend to utilize textbooks to assist their students. However, some studies (e.g. Hawkey, 2006), indicate that tests may encourage teachers to develop multiple materials rather than solely depending on textbooks. Wall and Alderson's (1993) Sri Lankan study states that a large group of teachers "believe they have to follow the textbook faithfully because the exam may test any of the content therein" (p. 63). Cheng's (1997) HKCEE (Hong Kong Certificate of Education Examination) and Fullilove's (1992, cited in Bailey, 1999) RUE studies reveal the booming market for publishing test-related materials. All these studies similarly find that most teachers heavily depend on textbooks.

Andrews, et al. (2002) also speak of the large role played by published materials in the Hong Kong classroom, citing a previous study by Andrews (1995)

in which the teacher respondents were found to spend an estimated two-thirds of class time working on exam-related published materials. Cheng suggests that a reason for this may be that the exam textbooks in Hong Kong not only provide information and activities but also suggested methods for teaching and suggested time allocations (1997). Read and Hayes (2003) note that in 90% of cases in their New Zealand IELTS study, exam preparation books were usually employed. One feature that the three Hong Kong studies have in common is that they investigate teachers' practices shortly after the introduction of revisions to a major exam. It would be interesting to see if similar findings emerged from a study conducted once the exam's contents and standards had become familiar to teachers; that is, how much were these results a fruit of uncertainty about the exam on the teachers' part? Alderson and Hamp Lyons (1996) indicate that at least in the situation they investigated, however, familiarity with the exam was not a variable, with many of the teachers, independently of their amount of experience of teaching towards the exam, making heavy use of exam materials. They suggest that one reason why teachers did this was that their negative attitude towards the exam discouraged them from creating their own materials.

Xiao (2002), on the other hand, has discovered that the test encourages the use of new textbooks and innovative teaching materials. Shohamy (1993) recounts a study that examined the impact of an Arabic test and found that it inspired the publication of new textbooks, which "have become, de facto, the new curriculum" (p. 10). However, Hawkey (2006), in his study of the impact on the Progeto Lingue (2000), shows that curricula designed to match the objectives of tests for Cambridge exams like KET, PET, and FCE, which emphasize communicative language approaches, may tend to encourage teachers to use additional materials instead of solely textbooks, from a variety of sources such as "cut-out photographs, self-designed spider games, information gap hand-outs, audio-cassettes, (and) wall charts" (p. 143).

Tests that emphasize a communicative approach, such as the HSC often elicit a heavy reliance on test-related materials by teachers. Progeto Lingue (2000) highlights a communicative approach, encourages the use of supplemental materials. This may be attributed to the purpose of test use. RUE and HKCEE are both high stakes and play a vital role in deciding students' academic futures. Because of this,

teachers devote more attention to assisting students to achieve high scores rather than learn real communication skills. It may be, then, that in the viewpoint of teachers, using test-related materials can assist them in doing their jobs better in terms of helping students receive better scores. Tests promote a boom of test related materials and thus influence what teachers teach in the classroom, but tests may also encourage teachers to use additional materials from a variety of sources. A teacher's choice of materials relies on a number of factors such as the purpose of the test and the availability of ready-made materials.

2.6.7 Washback on Lesson Contents

Learners follow a 'hidden' syllabus, that is, the contents driven by the contents of EFL examination. Many teachers, however, consistently skip over the listening lessons in their textbooks, because they know that listening will not be tested in the examination. A group teachers may 'do listening', but in a way that does not resemble the textbook designers' intentions. A few teachers cover the listening lessons if the type of question that students have to answer resembles an item type that might appear in the examination for reading. Most teachers in Bangladesh, particularly the higher secondary English school teachers, also admit they are influenced by the power of the public examinations. Thus, the status of their course is established by the importance of the teaching contents reflected on the entrance examinations.

There seems to be something of a mismatch between the attitudes of the teachers towards the contents of the learning package, and those of the students. The teachers clearly see the potential of the materials as a teaching package, containing relevant and worthwhile teaching activities, including but extending beyond test preparation. The students, on the other hand, are above all concerned with familiarising themselves with the format of the test, and seemed to be relatively little concerned with the learning strategies proposed, and the broader suggestions for improving performance.

In general, students demonstrate relatively little interest in the idea of using test preparation as an opportunity for language learning. Alderson and Wall (1993) conclude from their Sri Lanka study that 'the examination has had a demonstrable effect on the content of language lessons' (p, 126-27). This effect was that of the

narrowing of the curriculum to those areas most likely to be tested. This finding is similar to that of Lam (1994) who reports an emphasis in teaching on those parts of the exam carrying the most marks.

2.6.8 Washback on Learning Outcomes

Teaching to the test and test taking strategies might increase students' scores, but the score gains are not always statistically significant. Moreover, class instruction of exam-specific strategies and non-class instruction factors such as students' initial proficiency, personality, motivation, confidence, and exposure of environment all possibly contribute to a score gain. A test itself does not lead to various aspects of the perceived effects. It is rather mediating factors such as teachers' beliefs and educational backgrounds, students' individual differences (e.g. motivation, English proficiency), and purpose of test use that play essential roles in causing test effects.

It has been demonstrated that a test can result in all desired changes in teaching and learning. Wesche (1983), points out that when tests reflect the situations, content and purpose where learners will use the language, they are likely to improve motivation. Education is a complex phenomenon and there are many factors involved in bringing about changes, like the school environment, messages from administration, expectations of teachers and students, for example. Saif (2000) argues that an analysis of the needs and objectives of learners and educational systems should be carried out as a starting point for the research in washback.

Wesdorp (1982) finds there is no difference in students' writing in quality before and after the introduction of multiple-choice tests. Hughes (1988) reports that at a Turkish university, students' performance on the Michigan Test (a measure of English proficiency) increases after the introduction of a new test along with additional summer courses in English. Andrews et al. (2002) investigate the score comparisons that students receive on the UE (Use of English) oral exam in Hong Kong from 1993 to 1995. Students' scores have increased, but the score gain is not statistically significant. They claim that students' improved proficiency might have something to do with their "familiarization with the exam format, the rote-learning of exam specific strategies and formulaic phrases" (p. 220).

Elder and O'Loughlin (2003) examine the relationship between intensive English language study and band score gains on the IELTS and find there are great gains in listening, but no significant progress in reading skills. In Elder and O'Loughlin's study, a range of factors are linked to improving scores on tests, such as personality, motivation, confidence and exposure. Green (2006, 2007) finds students' initial scores instead of course length is a strong predictor of IELTS writing test score gain. In this sense, students' original proficiency plays a more important role in the resulting score gain than the time they spend in the test-preparatory course. Score gain washback, as concluded from the foregoing discussion, is a complicated issue. It is difficult to detect what causes or does not cause it. Further research needs to be conducted to determine whether students have made progress because the test motivates them to study harder or if other factors such as their original proficiency, personality, motivation, and exposure have more weight in explaining the outcome.

2.6.9. Strategies for Washback

It is seen that washback effects, on the one hand, may have potential for education, but on the other hand, may induce unexpected problems. The question is how to promote the intended washback of a test and minimise the possible counter productive reactions:

Firstly, the test must accurately reflect course objectives and the principles of mastering the knowledge need. This will lead teachers and learners to appropriate teaching and learning styles and enable beneficial washback to operate. If the test is at variance with the course objectives, it will require teachers to focus their teaching on the test alone and cause harmful washback.

Secondly, teachers, administrators and others involved should be trained and provided with information concerning the test, such as the aims, item type, scoring systems, specimen papers, etc. Competence and familiarity will help teachers and administrators to work properly toward the test, and limit misuse of test and its results. Next, test consequences play an important role in enabling either beneficial or harmful washback to operate. The more profound the consequence, the greater washback effect is. Educational settings would help to balance beneficial and

harmful washback in reducing test pressure toward teachers and students by appropriate continuous assessment.

   Additionally, parents or the public should be informed of the nature and the use of the test, as some political and social uses of test scores might induce unexpected harmful stresses on schools, teachers, and students (Smith, 1991). Furthermore, apart from the test itself there are many factors within a society, particularly the educational environment with its typical conditions all influence the behaviours of teachers and students. Nevertheless, to what extent these factors operate much depend on how they interact with each other in a specific circumstance. Although these factors and the test are interacted in a complex way, the following model (Figure 2.18) can be built to describe the interrelationship that enhances washback effects on teachers and students.

Figure 2.18: A model of the test development process (Saville, 2008)
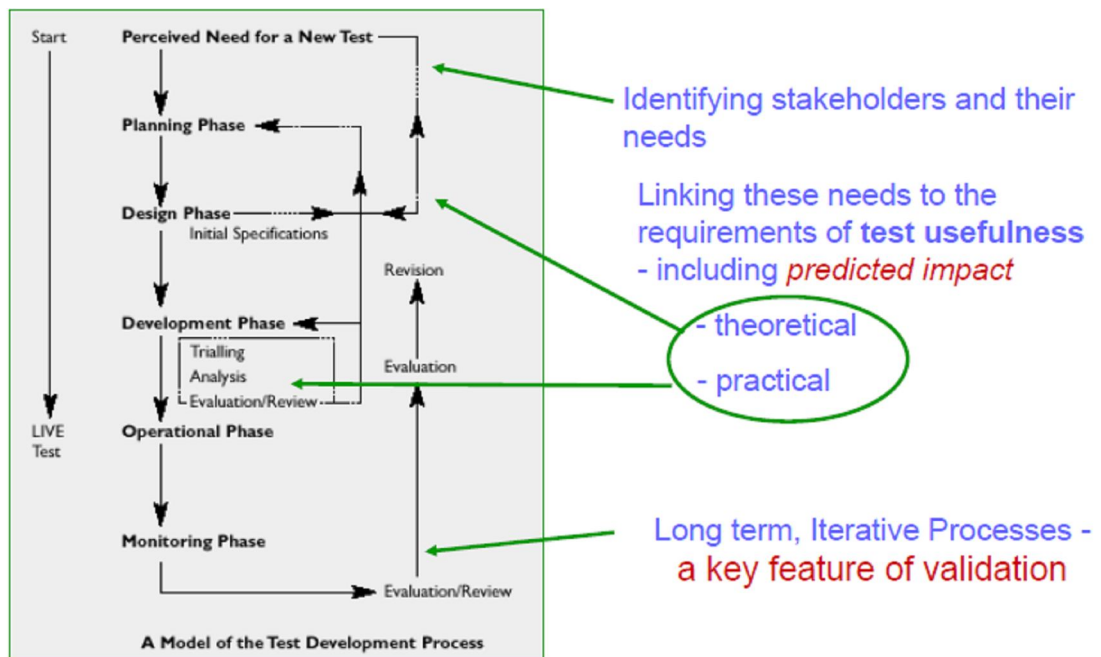
   Although precise descriptions of how tests have been reformed to promote washback are often lacking (Cheng, 2005; Wall, 2005), Hughes (2003) devotes a chapter to achieving beneficial washback. Brown (2000) summarises suggestions for the promotion of positive washback from Hughes (2003), Heyneman and Ransom (1990), Shohamy (1992), Kellaghan and Greaney (1992), Bailey (1996), and Wall (1996). Brown (2002) categorises these prescriptions as test design strategies, test

content strategies, logistical strategies and interpretation strategies. In the following outline, the present researcher attempts to summarize and organize the strategies proposed in the literature into four different categories that language educators can use to promote positive washback: test design strategies, test content strategies, logistical strategies, and interpretation strategies.

2.6.9.1 Test Design Strategies

A number of features of test design may be manipulated in efforts to improve instruction. These include item format (multiple-choice, short-answer question, extended response etc.), content (topics and skills), level of knowledge called for (retention, understanding or use), complexity (the number of content areas and their interrelationship), difficulty (easy or challenging), and discrimination (in terms of set standards of performance), referential source (criterion-referenced or norm-referenced), purpose (learner performance, curriculum evaluation, teacher evaluation) and type of items(proficiency, achievement or aptitude). Some specific strategies for designing a test to promote beneficial washback are:

1. sampling widely and unpredictably (Hughes, 1989),
2. designing tests to be criterion-referenced (Hughes, 1989; Wall, 1996),
3. designing the test to measure what the programs intend to teach (Bailey, 1996),
4. basing the test on sound theoretical principles (Bailey, 1996),
5. basing achievement tests on objectives (Hughes, 1989),
6. using direct testing (Hughes, 1989; Wall, 1996), and
7. fostering learner autonomy and self-assessment (Bailey, 1996).

2.6.9.2 Test Content Strategies

A number of researchers (e.g. Hughes, 1989; Heyneman and Ransom, 1990; Bailey, 1996) have suggested some test content strategies to balance beneficial and harmful washback in reducing test pressure toward teachers and students by appropriate continuous assessment:

1. testing the abilities whose development you want to encourage (Hughes, 1989)

2. using more open-ended items (as opposed to selected-response items like multiple choice) (Heyneman and Ransom, 1990)

3. making examinations reflect the full curriculum, not merely a limited aspect of it (Kellaghan and Gleaney, 1992)

4. assessing higher-order cognitive skills to ensure they are taught (Heyneman and Ransom, 1990; Kellaghan and Greaney, 1992)

5. using a variety of examination formats, including written, oral, aural, and practical (Kellaghan and Greaney, 1992)

6. not limiting skills to be tested to academic areas (they should also relate to out-of-school tasks) (Kellaghan and Greaney, 1992), and

7. using authentic tasks and texts (Bailey, 1996; Wall, 1996).

2.6.9.3 Logistical Strategies

The outcome of test use involves the collaborative efforts made by various stakeholders such as teachers, students, policy-makers and test-developers. Some logistical strategies as suggested by researchers are:

1. insuring that test-takers, teachers, administrators, curriculum designers understand the purpose of the test (Bailey, 1996; Hughes, 1989)

2. making sure language learning goals are clear (Bailey, 1996)

3. where necessary, providing assistance to teachers to help them understand the tests (Hughes, 1989)

4. providing feedback to teachers and others so that meaningful change can be effected (Heyneman and Ransom, 1990; Shohamy, 1992)

5. providing detailed and timely feedback to schools on levels of pupils' performance and areas of difficulty in public examinations (Kellaghan and Greaney, 1992)

6. making sure teachers and administrators are involved in different phases of the testing process because they are the people who will have to make changes (Shohamy, 1992)

7. providing detailed score reporting (Bailey, 1996)

2.6.9.4 Interpretation Strategies

Hughes (1989), Heyneman and Ransom (1990), Shohamy (1992), Kellaghan and Greaney (1992), Bailey (1996), and Wall (1996) all have provided lists of strategies for using the washback effect to positively influence language teaching. For more extensive discussion of these lists (Brown, 1997, 2000) some of the interpretation strategies are listed below:

1. making sure exam results are believable, credible, and fair to test takers and score users (Bailey, 1996),
2. considering factors other than teaching effort in evaluating published examination results and national rankings (Kellaghan and Greaney, 1992),
3. conducting predictive validity studies of public examinations (Kellaghan and Greaney, 1992),
4. improving the professional competence of examination authorities, especially in test design (Kellaghan and Greaney, 1992),
5. insuring that each examination board has a research capacity (Kellaghan and Greaney, 1992),
6. having testing authorities work closely with curriculum organizations and with educational administrators (Kellaghan and Greaney, 1992), and
7. developing regional professional networks to initiate exchange programs and to share common interests and concerns (Kellaghan and Greaney, 1992).

Test design and content strategies are more closely identified with washback direction, while logistical issues are more closely identified with washback intensity. Interpretation strategies may be viewed as indirect, policy-level means of ensuring standards of test design and logistical provision while the test design and content strategies relate most closely to Chapman and Snyder's (2000) test description categories of format, content, complexity and referential source. The Communicative approach to EFL teaching and learning has become increasingly accepted in schools and colleges in Bangladesh in recent years. A great deal of time and energy has been expended in developing materials and techniques to help achieve what has been termed *Communicative Competence.* Communicative language testing is intended to provide the tester with information about the testee's ability to perform in the target language in certain context-specific tasks. Strategies of language testing should be designed in such a manner that it can generate positive washback on language learning.

2.6.10 Washback Stakeholders

Washback is the result of a partnership between all *direct a*nd *indirect* participants whose relationships involve a constant multi-directional interplay. It has long been believed that tests directly influence educational processes in various ways. One common assumption is that teachers will be influenced by the knowledge that their students are planning to take a certain test and will adapt their teaching methodology and lesson content to reflect the test's demands. The term 'backwash' has been used to refer to the way a test affects teaching materials and classroom management (Hughes 1989), although within the applied linguistics and language testing community the term 'washback' is more widely used today (Weir 1990; Alderson and Wall 1993; Alderson 2004). Taylor (2000, p. 2), building upon a model proposed by Rea-Dickins (1997) identified at least 5 stakeholder categories: learners, teachers, parents, government and official bodies, and the marketplace, offers a more detailed conceptualisation in order to illustrate the wider societal effects of a test (i.e. test impact). Figure 2.19 illustrates how different stakeholders are involved in testing and tests scores:

Figure 2.19: Stakeholders in the testing community (UCLES, 2009)
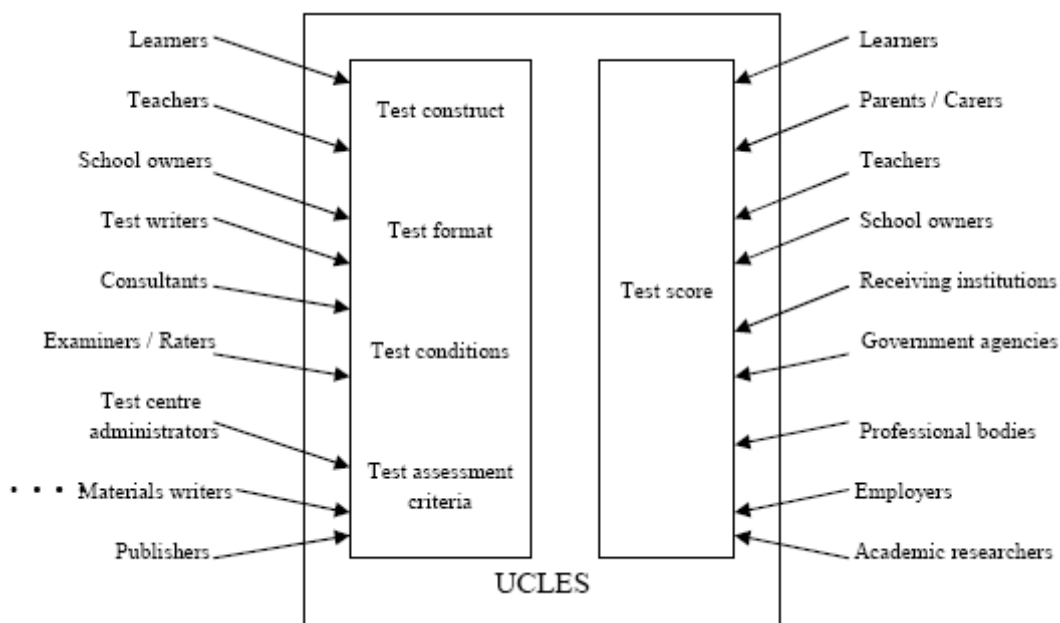
The above model provides a useful illustration of the fact that a test can have impact upon the various stakeholders involved, at different points in the testing process: Some of the stakeholders listed above (e.g. examiners and materials writers) are likely to have more interest in the 'front end' of a test, i.e. the test assessment criteria or test format. Others may see their stake as being primarily concerned with the test score. Some stakeholders, such as learners and teachers, will naturally have an interest in all aspects of the test.

As Pearson (1988) remarks, 'There is an explicit intention to use tests, including public examinations, as levers which will persuade teachers and learners to pay serious attention to communicative skills and to teaching learning activities that are more likely to be helpful in the development of such skills.'(p, 33).The past ten years have seen a growing awareness that testing can have consequences beyond just the classroom. Tests and test results have a significant impact on the career or life chances of individual test takers (e.g. access to educational/employment opportunities). They also impact on educational systems and on society more widely: for example, test results are used to make decisions about school curriculum planning, immigration policy, or professional registration for doctors; and the growth of a test may lead publishers and institutions to produce test preparation materials and run test preparation courses. The term 'impact' is generally used to describe these consequences of tests (Bachman 1990; Bachman and Palmer 1996). Some language testers consider washback as one dimension of impact, describing effects on the educational context (Hamp-Lyons 1997); others see washback and impact as separate concepts relating respectively to 'micro' and 'macro' effects within society.

It is worth mentioning fact that a test can have impact upon the various stakeholders involved, at different points in the testing process. Some of the stakeholders, e.g. examiners and materials writers are likely to have more interest in the 'front end' of a test, i.e. the test assessment criteria or test format. Others may see their stake as being primarily concerned with the test score. Some stakeholders, such as learners and teachers, will naturally have an interest in all aspects of the test.

Taylor (2000) offers a detailed conceptualisation in order to illustrate the wider societal effects of a test, building upon a washback model proposed by Rea-Dickins. Testing tends to induce consequences for its stakeholders. It is well known

in the field of education that there is a set of relationships, intended and unintended, positive and negative, between testing, teaching and learning. Impact refers to the effects that a test may have on individuals, policies or practices, within the classroom, the school, the educational system or society as a whole. Washback (also known as backwash) refers more frequently to the effects of tests on teaching and learning. Primarily, the effects of testing have been associated with test validity (consequential validity) and with test scores and score-based inferences to test use and the consequences of test use. Figure 2.20 displays the relations of stakeholders to testing and test scores:

Figure 2.20: Saville's stakeholders of macro-level washback (2008)
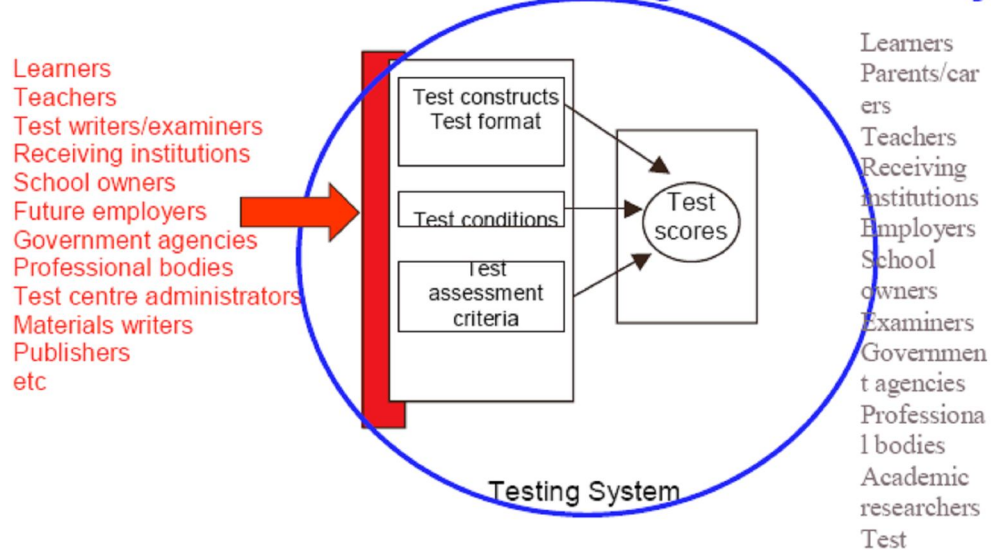
This presentation (Figure 2.20) will focus on first delineating impact, washback and consequences of large-scale testing and then report a series of empirical studies to illustrate the methodology used to research such a phenomenon in education. Washback research on other participants influenced by program washback is less widely developed than the research on washback effects on language learners and teachers. It is found that teachers are the most frequently studied participants in washback processes. However, many other people are also involved in language testing washback. The comparative dearth of empirical findings on students suggests that research is needed about how tests actually influence language learners' behavior and attitudes.

The research on other parties who try to create, or are influenced by, program washback is less widely developed than the research on language learners and teachers. The other participants can include test developers (Andrews, 1994b; Andrews & Fullilove, 1994), teacher educators and curriculum planners (Andrews & Fullilove, 1994), teacher advisors (Wall & Alderson, 1993), principals and other administrators (Fullilove, 1992; Hughes, 1993; Shohamy, 1993b; Shohamy, Donitsa-Schmidt, & Ferman, 1996), language inspectors (Shohamy, Donitsa-Schmidt, & Ferman, 1996), end-users (Andrews & Fullilove, 1994), materials developers and publishers (Cheng, 1997; Hughes, 1993), and even parents (Andrews, 1994a; Cheng, 1997; Fullilove, 1992; Ingulsrud, 1994; Shohamy, Donitsa-Schmidt, & Ferman,1996).

A repeated theme found in the literature on these other participants, particularly test designers and policy makers, is the dynamic tension between (1) the intended positive washback in implementing new or revised exams and (2) how that impact is realized in classroom practices. Andrews and Fullilove (1994, pp. 57-58) assert that in cases where new or revised tests have a negative washback effect, the reforms in language teaching proposed by teacher educators and curriculum planners have been undermined by the conflicting message implicit in the tests, especially in those countries where examinations are highly important and yet where the examination format has been particularly resistant to change.

The data in their study included structured interviews with teachers and with inspectors of ASL (ibid., p. 302). The authors found in their interviews that the inspectors were aware of high test anxiety (among both teachers and students) in previous years' tests, but that the test anxiety had decreased and that some teachers did not even administer the test. Others treated it as a quiz that required no preparation. However, Shohamy et al. stated the inspectors felt that it is essential that the test continue to be administered as they believe that there would be a major and significant drop in the level of Arabic proficiency in the country were the test to be cancelled. Moreover, the Inspectorate claims that there would be a decrease in the number of students studying Arabic since the test promotes the status of Arabic as perceived by teachers, students and parents. (ibid.). This finding illustrates the disparate views held by the inspectors, on the one hand, and the students and teachers of Arabic on the other.

When Shohamy et al. (1996) interviewed the inspectors associated with the high-stakes EFL exam, they found that "the Inspectors claim that the introduction of the oral test has had a very positive educational impact and the washback on teaching has been tremendous" (ibid., p. 312). The inspectors also feel that the test has successfully promoted learning, particularly of oral skills. They believe that "were the oral exam to be cancelled, teachers would cease teaching oral proficiency" (ibid.). In other words, in both cases, the inspectors of the Arabic and English exams see their respective tests as "necessary, important and effective" (ibid., p. 313).

However, Shohamy et al. point out that this position "is in contrast to how teachers and students perceive the test" (ibid.) and that in general "unlike teachers and students, the bureaucrats portray a much more positive picture" (ibid.). Another set of participants who may be influenced by or try to utilize washback is the "end-users"- that is, people who, in the future of the language learners, will in some way benefit from their target language proficiency. (In the English for Specific Purposes [ESP] literature, the students' future employers are often the end-users.). In this case the tertiary institutions may be seen as the "end-users" who have a stake in the product of secondary school English teaching- that is, the future university students' ability to use oral English.

Finally, parents are occasionally included in research on washback phenomena. Andrews (1994a) notes that there is "widespread acceptance of the assertion that tests, especially public examinations, exert an influence on teachers, learners and parents (p. 45). Anxious parents take their tiny 'scholars' to pre-kindergarten interviews to gain admission to choice places even on this lowest rung of the educational ladder. However, there is relatively little research that documents the parents' own perceptions of language testing washback. The studies that document parents' ideas typically do so through the students' perspective. However, many other people are also involved in language testing washback. The comparative dearth of empirical findings on students suggests that more research is needed about how tests actually influence second language learners' behavior and attitudes.

2.7 Implication of the Theoretical Perspectives for Washback Study

The theoretical perspectives as well as the research evidence presented above cast new light on the recurring themes that have been previously discussed. The framework will help future researchers and other readers conceptualize the whole teaching process. It is beyond doubt that Woods (1996) as well others (Ernest, 1989; Fang, 1996; Nunan, 1999; Reagan & Osborn, 2002; Richards, 2008; Samuelowicz & Bain, 1992, 2001; Shulman, 1987; Thompson, 1992; Williams & Burden, 1997; Yates & Muchisky, 2003) has a wide range of implications for our understanding of the role of the different factors in washback. Not only do they provide researchers and readers with a comprehensive understanding of the complex reality of innovation, but they also offer all a different way of thinking about notions such as teacher beliefs, knowledge, and experience (BKE) and their connection to teaching and learning EFL. Moreover, the interdisciplinary theoretical framework provides me with a broad set of conceptual tools for systematic investigations of teacher thinking and its relationship to teacher classroom practice.

Specifically, this theoretical framework can inform the present study at least from different bases. One example drawn from washback studies to illustrate is that a number of researchers have found it hard to make weighty claims, and thus made only tentative ones. It appears that these researchers may have failed to take into account the developmental characteristic of change. Since the research focus of the majority of studies is short-term, no conclusions can be drawn about long-term washback effects. The theoretical frameworks above have also offered us enlightening insights into how to look at and cope with conflicts, constraints, differences and discrepancies that have emerged from innovation. The implementation of educational reforms, including testing reforms calls for the conceptual change in teachers' pedagogical beliefs, teachers' perspectives interwoven characteristic of teachers' dynamic aspects of teachers' beliefs, assumptions, knowledge, etc.

This, with respect to washback research, can be interpreted to mean that in order for teachers to change their perceptions of tests, they need to change their perceptions of teaching and learning, and their perceptions of language as well, for all these beliefs are intrinsically interwoven. To be specific, teachers' and students' beliefs of tests are likely to correspond to their beliefs of language teaching and

learning. Meanwhile, their beliefs of language teaching and learning are likely to follow their conceptions of what is meant by learning as well as their beliefs what language is. Here, the relationship between beliefs of language teaching and beliefs of language learning is also interactive and interconnected. All these beliefs and attitudes are crucial in the sense that they may not only influence but also affect the way they interpret and react to washback. Such a basis not only helps to clarify the complexity of the innovation process, but also helps to improve further innovation endeavors. Therefore, there is a need to apply these insights to washback research.

It examined the research on washback in language education and general education to clarify and summarise some basic concepts and theoretical perspectives related to the washback phenomenon. It provides a general conceptual framework in an attempt to highlight the overlapping patterns and themes that have emerged through the lens of this framework. It allows the present researcher to document and interpret the washback phenomenon of the HSC examination on the EFL education in Bangladesh

2.8 Conclusion

The theoretical framework discussed above indicates that during the last decade the interest in washback has not only grown, but it has also focused on what forms washback takes, indications of its appearance in specific environments and its influence on participants, processes, and the associated products. The theoretical framework of washback has produced some evidence that it exists. Such research also highlights the complexity of the washback phenomenon and some of the difficulties involved in designing, implementing and interpreting research in this area. There are concerns that the introduction or changes to a test may create a negative washback effect, particularly in the case of high stakes tests such as the HSC examination in EFL in Bangladesh. However, whether the influences of testing on teaching and learning are positive or negative is still debatable and needs to be studied further.

In this chapter, an attempt has been made to clarify the definition, scope, and function of washback for the purpose of this study. Washback is at the heart of the intricate relationship between testing, teaching and learning. It also illustrates the impact and power of tests on teaching and learning in educational contexts. A large number of studies have dealt with the phenomenon of washback from different

105

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

perspectives and at multiple levels. There have, however, been few empirical analyses that have investigated how the washback phenomenon actually happens in the classroom. There have been even fewer research studies that have considered washback at both the macro and micro levels, particularly in language education.

Discussions in this chapter have reviewed a number of studies in searching for the meaning and mechanism of the function of washback, including Alderson and Wall's (1993) 15 washback hypotheses, and the models of the mechanism of washback as a phenomenon of change in teaching and learning. These models have helped the current study to determine the nature of washback, and how washback works in educational contexts and they seem particularly appropriate, as the general aim of the present study is to examine and understand the function of washback on teaching and learning English at the HSC level. Together, those models have allowed the present researcher to formulate the central issues that will be explored in the current study. By combining the models, possible washback effects in an area or in a number of areas of teaching and learning affected by tests can be investigated. Accordingly, a study of the effects of washback needs to draw on curriculum and innovation models and explore the phenomenon within a multidimensional context. A working framework for this study has been built on this basis, as presented in Figure 4.1 (Chapter Four). A model of washback will be proposed in order to describe explicitly possible catalysts between assessment and washback effects, based on the findings of the research.

In the light of issues raised in previous studies, it is clear that a study looking at what and how the HSC examination influenced teaching and learning at the HSC level in Bangladesh would need to focus on the following dimensions: what possible areas of English teaching and learning have been affected by the tests; how different levels of stakeholders within the Bangladesh educational system have reacted when washback occurred; defining the interrelationship between who changes what, how, when, where, and why. The chapter has discussed the new meaning of and insights to the research on washback. After the brief introduction provided in Chapter One to the general context of the study, this chapter has presented a broad set of theoretical tools outlined from multiple sources. The next chapter presents a literature review of washback studies on the EFL/ESL teaching, learning and testing. An extensive discussion of studies in other research areas that influence and shape the present study is also highlighted in the following chapter.

This chapter presents a review of different bodies of literature relevant to the present study. Its purpose is to gain insight into the complex dimensions of washback, and illuminate the vital role that washback plays in ESL and EFL education. That is it contains an overview of the advances in washback research over the past two decades. It focuses on the importance and objectives of literature review; and it finally summarises a number of relevant research studies on different domains of language education that washback affects. Specifically, it draws on ideas from language education, general education, psychology and other innovation research to see whether insights can be gained into the patterns and themes that have recurred in washback research.

3.1 Overview of the Advances in Washback Research

Review of literature surveys dissertations, scholarly articles, books and other sources (e.g. conference proceedings, etc.) relevant to a particular issue, area of research, or theory, providing a description, summary, and critical evaluation of each work. A literature review is a body of text that aims to review the critical points of current knowledge and or methodological approaches on a particular topic. Literature reviews are secondary sources, and as such, do not report any new or original experimental idea. Most often associated with academic-oriented literature, such as theses, a literature review usually precedes a research proposal and results section. It brings the reader up to date with current literature on a topic and forms the basis for another goal, such as future research that may be needed in the area.

A well-structured literature review is characterized by a logical flow of ideas; current and relevant references with consistent, appropriate referencing style; proper use of terminology; and an unbiased and comprehensive view of the previous research on the topic. For the present study, the researcher has collected information from various sources: a good number of books, a number of dissertations and journal articles, and information from internet sources. This chapter incorporates a critical

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

107

review of the relevant literature with particular attention on washback definitions, its connection to impact, positive and negative connotations, models of test washback and it presents an overview of some major washback studies.

It is worth mentioning that language testing researchers have embraced the call from Alderson and Wall (1993) for more intensive research on washback. During the last two decades, the researchers have accomplished a substantial volume of research on this topic (e.g. Alderson & Hamp-Lyons, 1996; Cheng, 1997, 2004; Cheng & Qi, 2006; Green, 2006, 2007; Muñoz & Álvarez, 2010; Qi, 2004, 2007; Saif, 2006; Shih, 2007; Shohamy, 1993; Shohamy et.,1996; Tan, 2008; Turner, 2001, 2005, 2008, 2009; Wall, 1996, 1999; Wall & Alderson, 1993; Wall & Horák, 2008; Watanabe, 2004b). In recent years, researchers have been making significant inroads into investigating this phenomenon in different social and educational contexts. As a result, the definition as well as the nature and scope of washback have been extensively discussed, and a number of different perspectives have emerged in language testing and ELT research areas. The reviews, taken together, constitute a general framework for looking at the research topic in this study.

3.2 Research on Washback in Applied Linguistics

Despite the strong link between testing, teaching and learning discussed in the field of education, the assertion that a test influences what teachers and students do in the classroom is often based on anecdotal evidence, and did not receive much attention from researchers until the early 1990s in the field of applied linguistics (Andrews, 2004; Bailey, 1996; Wigglesworth & Elder, 1996; Wall, 2000; Watanabe, 1996). Between 1980 and 1990, little empirical research had been carried out to investigate the washback effect of examinations either in the field of general education or in the field of language education.

Although Alderson (1986) recognises the potential use of language tests as a tool to bring about positive effects on language teaching and learning about two decades ago, it took almost another 10 years for the concept of tests influencing teaching and learning to become an established research topic. McNamara (2000) argues that this is because applied linguistics researchers tend to focus heavily on investigating individuals' language skills and abilities, rather than on the

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

108

consequences of tests. Elder and Wigglesworth (1996) also point out that the concept of tests influencing teaching and learning is under-researched probably because the huge number of variables involved have made it very difficult for researchers to identify a causal relationship between the test and what goes on in the classroom.

Figure 3.1: The three stages of effective literature review process (Levy & Ellis, 2006)
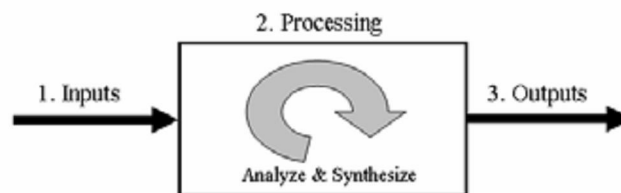
Though a good number of washback studies have been carried out during recent years, the washback effect is still to be adequately defined and analysed. While there is consensus that washback incorporates the effects of tests on teaching and learning, researchers have not agreed on what washback is, what it might look like, or how it works. There have only been a limited number of washback studies, and invariably, researchers call for further investigations that would establish what washback is and even whether it exists. This chapter incorporates a critical review of the relevant literature. It also summarises washback related research with emphasis on the washback effects and impact of the test.

The present researcher reviewed only those works that were directly relevant to the present study. The reviewed literature mainly includes scholarly books, dissertations, research articles, monographs, and periodicals for the development of insights into the present study. The present researcher reviewed the relevant literature for many other reasons, such as research methods and techniques, new ideas and approaches, what needs to be done, relationships between ideas and practices, correlations, contradictions between the findings of the present study and those of the reviewing studies, etc.

This review of the literature is presented in three sections: (1) the first section presents the studies carried out from 1982 to 1999; (2) the second section focuses on the washback research conducted from 2000 to 2005; and (3) the third section includes the studies carried out from 2006 to date.

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

109

3.2.1 Washback Studies from 1982 to 1999

Kellaghan, et al. (1982) conducted a study on "The effects of standardized testing" which had extensive potentials for the future researchers. So far, it is considered the first study on washback. They studied the educational and psychological effects of the introduction of standardized achievement/ability testing in elementary schools in Ireland. The study by Kellaghan et al. is considerd to be of high quality. Kellaghan et al. observed that teachers in Irish primary schools were quite biased in the evaluation of their students at the time of their study. They speculated that the reason such bias existed was due to the lack of standardized testing in Ireland.

Wesdorp (1982) carried out a research on "Backwash effects of language testing in primary and secondary education in Netherlands" which investigated the validity of objections to the use of multiple-choice tests for the assessment of both first and foreign language education. The results did not support the assumed negative washback effects. One of the assumptions, for example, was that the skills that could not be tested by multiple-choice questions would not be taught any more in primary schools. Differences between the teachers' activities in schools with and without a multiple-choice final test were insignificant. The results did not show any changes in the students' study habits either. On the whole, the study revealed much less negative washback than had originally been assumed. However, it is not clear what kind of tests had been in effect before the introduction of multiple-choice tests and how different the tests measuring first and second language education were. It could be that the old test methods (e.g., direct/indirectness, discrete-pint/integrative approach) and content were so similar to those of multiple-choice tests that even after the introduction of the new technique teachers and learners didn't feel any need to change their attitudes towards the tests.

Hughes (1989) described a project conducted in a non-English speaking country, at a Turkish English-medium university. Before the study started, undergraduate students used to enter academic programmes after spending a year of intensive English study, yet they demonstrated a very low level of English proficiency. As a result, the university decided to establish a screening device to determine which students could continue with their studies and which students would have to leave the university. A new test was developed based on the English

study skills needs of freshman students (e.g., reading, note-taking, etc.) which included tasks similar to those they would have to perform as undergraduates. Hughes (1989) reported that the introduction of this test in place of the old multiple-choice test immediately affected teaching.

Khaniya (1990), in a study in Nepal, attempted to study washback by designing a new communicative English language proficiency test and comparing it with the traditional SLC (School Leaving Certificate). According to Khaniya, the SLC had important consequences for the future of the students since it was a factor in the selection of university and job candidates. Consequently, students, teachers and parents were very much concern with its results. As Khaniya descried, SLC required students to memorise texts and answered to questions since many of the test questions and texts were taken directly in the textbooks. In such a situation, the exam would definitely have some sort of control over the course, but he did not explain how teachers actually taught to the exam, what and how students learned, and so on. He gave his new test to three different groups of students at the beginning and at the end of grade 10 when students were preparing for the SLC. Based on final results, Khaniya reported that while the differences between students' performance (in English-medium schools) before the introduction of the new test was not significant, at the end of the year those with an emphasis on skills improved their performance on the new exam while the students whose program emphasized SLC performed poorly- Khaniya claimed that this is because of the SLC examination teaching going on in exam-emphasizing schools, due to the negative washback of the SLC test. He argued that the fact that the third group of students (in Nepalese-medium schools) also performed poorly at the end of the year further supported this claim.

Li (1990) conducted a research on the Matriculation English Test (MET). It is the English language test for entrance into all universities in China, which has been the subject of several washback studies. It is a standardised, norm-referenced proficiency test, which in 1990 had an annual test population of 3 million. Li documented the evidence for washback four years after the MET had been introduced. Data was collected through the analysis of test results and their comparison with other tests. A study of student writing was also carried out. A number of 229 teachers completed the questionnaire. Students were also questioned.

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

111

Their typical response was that the good thing about MET was that they did not need to memorise in order to prepare for it, a major departure from the usual tests they sat. The study recorded the washback effects of the new test over a five-year period and found it encouraged the use of new textbooks and innovative materials. Although Li noted that some of the changes the research had uncovered were not all that significant in terms of encouraging high school teachers to change their teaching methods, she was hopeful that there would gradually be a marked and more persistent change over time.

Smith (1991) reported on two qualitative studies which investigated the effect of tests on teachers and classrooms. Data from interviews revealed that the publication of test results induced feelings of fear, guilt, shame, embarrassment, and anger in teachers, and the determination to do what was necessary to avoid such feeling in the future. Teachers believed that test scores were used against them, despite the perceived invalidity of the scores, and they also believed that testing had severe emotional impact on young children. From classroom observation, it was concluded that testing programmes substantially reduced the time available for instruction and narrowed the curriculum and modes of instruction. Smith reported that there were two different reactions to this "narrowing of the curriculum". One was accommodation by teachers, who discarded what was not going to be tested, and taught towards the test, other was one of resistance, exemplified by one teacher, he said that he knew what was on the test, but he felt that children should keep up with current events and trace the history behind what was happening then, so they were going to spend march doing that. This suggests that washback phenomenon is not quite as simple as is at times made out.

Alderson and Wall's (1993) "Does Washback Exist?" was a great manuscript on washback. It was considered as a milestone in the field of washback study. In their manuscript, the concept of washback, or backwash, defined as the influence of testing on instruction, was discussed with relation to second language teaching and testing. It is an empirical research, where they coined 15 hypotheses on the washback. Much of the literature on this subject had been speculative rather than empirically based. They were the first scholars to suggest that the washback effects of language tests were not as straightforward as had been assumed. It was Alderson and Wall who pointed out the problematic nature of the concept of washback and the

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

112

need for carefully designed research. In their article 'Does Washback Exist?' they questioned existing notions of washback and proposed a series of washback hypotheses. These hypotheses are potentially playing a role in the washback effect and must therefore be considered in any investigation.

They suggested that tests were commonly considered to be powerful determiners of what happens in the classroom, the concept of washback was not well defined. The first part of the discussion focused on the concept, including several different interpretations of the phenomenon. It was found to be a far more complex topic than suggested by the basic washback hypothesis, which was also discussed and outlined. The literature on education in general was then reviewed for additional information on the issues involved. Very little research was found that directly related to the subject, but several studies were highlighted. Following this, empirical research on language testing was consulted for further insight. Studies in Turkey, the Netherlands, and Nepal were discussed. Finally, areas for additional research were proposed, including further definition of washback, motivation and performance, the role of educational setting, research methodology, learner perceptions, and explanatory factors.

Alderson and Wall (1993) carried out longitudinal study on "Examining Washback: The Sri Lankan Impact Study" in Sri Lanka concerning the effects of second language tests, specifically the O-Level examination in English as a Second Language; on classroom language instruction is reported. This was the landmark research on washback. Their study investigated the phenomenon of washback or backwash, the influence of testing on instruction. Their study was cited as the only known research investigating washback in language education through consecutive classroom observation. The study was conducted at the secondary school level, and combined classroom observation with data from interviews, questionnaire responses, and test analyses to determine whether washback existed, to what degree it operated, and whether it was a positive or negative force in this educational context.

This long-term impact study was jointly conducted by a research team over a period of two years. It differed from other studies in that it was the most comprehensive and thorough study that had ever been conducted in this research area. The entire study was composed of several sub-projects: a baseline study, questionnaires to teachers and teacher advisers, teacher interviews (group),

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

113

document and material analyses (especially tests), and, most importantly, a two-year observation programme. It is worth noting that the research team (7 Sri Lankan teachers) conducted six rounds of classroom observations in a total of 49 schools across the country. The findings gave background information on the project; discussed the characteristics of positive and negative washback in terms of instructional content, instructional methods, and techniques, and assessment and presents the results of two rounds of classroom observation. The study concluded that washback occurred in both positive and negative forms, to some degree, in teaching content, but not in methodology. Existence of washback, both positive and negative, on the way teachers and local education officers design tests was also found. They recommended further research on this field.

The study of Herman and Golan (1993) looked at the effects of standardized test on teaching and learning processes in upper elementary classrooms in eleven districts in nine states. The study investigated test washback in a holistic way by looking at the self-reported influences of tests within classroom settings as well as on policy-makers, which had contributed a great deal of understanding washback from a macro point of view. Data was collected from 341 teachers for their study. The study revealed the pressure that teachers felt to improve test scores and the amount of time teachers spent on test preparation. Results indicated that standardized testing had considerable effects, and that teachers felt considerable pressure to improve student scores (SLD). The findings reported that over 50% of the teachers admitted that they would give substantial attention to mandated tests in their instructional planning and delivery. In devising their syllabi for instruction, they would look at prior tests to assure that they covered the subject matter of the test or test objectives.

Stephens et al. (1995) conducted a research using a case-study approach. The study sought to describe what assessment looked like in four school districts (two schools per district, two classrooms per school). Interviews were conducted with students, parents, teachers, principals, and central office staff to understand assessment from multiple perspectives. Teachers were interviewed prior to and after three half-days of observation to understand assessment as part of classroom practice. Results indicated that the meanings of particular concepts, such as assessment, curriculum, and accountability, varied significantly across districts. The

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

114

salient relationship was not the one between assessment and instruction, but rather the relationship of each of these to the decision-making model of the district. Generally, when assessment-as-test did appear to drive instruction, this relationship seemed to be an artifact of a model in which individuals ceded authority for decision making to outsiders. When assessment as test did not appear to drive instruction, this relationship seemed to represent a model in which individuals maintained the authority to make decisions within the framework of their individual and collective philosophies. Findings suggested that assessment as test did not necessarily drive instruction, and that when assessment as test did drive instruction, it did not drive it in a way that might be considered good instruction.

Alderson and Hamp-Lyons (1996) studied a longitudinal study which examined how washback of public exams impacted English teaching in Sri Lanka. Their study provided insights into the relationship between teachers' perceptions of teaching contents and public examinations. Two points in Alderson and Hamp-Lyons's study were particularly strong. First, they incorporated an observational component in their study rather than relying solely on self-reports. Second, they used laughter as one barometer of the classroom atmosphere. A second limitation of Alderson and Hamp-Lyons's study was their choice of participants.

Alderson and Hamp-Lyons pointed out that the TOEFL affected both what and how teachers taught, but the effect differed considerably from teacher to teacher. It would be worthwhile to determine whether those effects were similar among teachers with comparable backgrounds. A third concern about the study by Alderson and Hamp-Lyons was that they dealt with washback primarily from teachers' perspectives, barely addressing students' points of view. They commented, to better understand how washback occurred within the classroom, researchers needed to investigate changes in students' motivations, learning styles, and learning strategies. One final concern about Alderson and Hamp-Lyons's study was that they did not make it clear what - if any - student score gains occurred.

Shohamy et al. (1996) examined the impact of national tests of Arabic as a Second Language (ASL) and English as a Foreign Language (EFL) in Israel. They explored different washback patterns among teachers, students, and inspectors in terms of how these tests influenced classroom activities, time allotment, teaching materials, perceptions of prestige, and the overall enhancement of learning.

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

115

Regarding the EFL test, oral teaching activities were progressively introduced. As a consequence the amount of instruction time for oral activities increased, new courseware was brought in, awareness of the test increased, and the subject matter's status in the school substantially rose. In contrast, the ASL's impact in those areas declined to the point of insubstantiality. Nevertheless, the bureaucrats believed both tests had reached their objectives without any need for teacher training or curricular revision.

Their research found that teachers were motivated to implement activities to promote their students' skills for the test. A change of how teachers would evaluate their students due to the influence of public exams was found in an empirical study regarding the new EFL test in Israel. According to Shohamy et al. (1996), "the rating scales which measure accuracy and fluency will be changed slightly and a new scale of task orientation will be added". The study concluded that washback changes with time because of factors such as language status and test uses.

Cheng, L. (1997) conducted a study on **"How Does Washback Influence Teaching? Implications for Hong Kong"** to investigate whether or not any washback effect of the revised Hong Kong Certificate of Education Examination in English (HKCEE) by the HKEA could be observed in the teaching of English in Hong Kong secondary schools. The aim of the study was to observe how the whole education system would react in the context of the change in its assessment practice and to attempt to discover the implications of the washback effect on the teaching of English in Hong Kong secondary schools.

The HKCEE was a public examination taken by the majority of secondary students at the end of the fifth year of their secondary school. Two separate syllabuses, namely the examination syllabus by HKEA and the teaching syllabus by the CDC (Curriculum Development Council) coexisted in Hong Kong secondary schools. Her research employed various methodological techniques such as questionnaires, interviews, and classroom observations, which were based on an in-depth case study approach to sampled schools in Hong Kong. She conducted the study among the 42 students, and 48 teachers. The study took place from January 1994 to November 1996 and consisted of three phases.

Although the Hong Kong Examinations Authority intended to create a positive washback effect through the innovation, Cheng's findings indicated that

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

116

changes occurred mainly at a superficial level: the content of teaching and the materials used changed rapidly but there was not much evidence of fundamental changes in teaching practices and student learning. When teachers were asked about their reaction to the new examination, 37% of them were sceptical about the changes, 29% were neutral and another 2 1% welcomed or enthusiastically endorsed the changes, with 13% of teachers not responding to the question.

It was found that 84% of the teachers commented that they would change their teaching methodology as a result of the introduction of the 1996 HKCEE. While 66% of the teachers mentioned that the proposed changes in the 1996 examination syllabus might not contradict their present teaching methodology, 68% of teachers felt the new examination would add pressure to their teaching. It was that 61% of the respondents stated that the selection of particular textbooks was made by teachers jointly. As to general lesson arrangement, decisions were made by teachers according to 60% of the respondents and panel chairs according to 29% of the respondents. When teachers were asked how they carried out language skill training in class, they replied that 61% of the English lessons were arranged for the purpose of teaching separate skills such as listening, reading or grammar usage. Only 5% of the lessons were arranged on the basis of integrated skills including listening, speaking, reading, and writing.

The Findings indicated that the washback effect worked quickly and efficiently to bring about changes in teaching materials, largely due to the commercial characteristics of Hong Kong society, but somewhat slowly, reluctantly, and with difficulty in the methodology that teachers employed. The study suggested that teaching content had so far received the most intensive washback effects, although washback effects had also been observed in teachers' attitudes and behaviors and in the English curriculum.

Watanabe, Y. (1996) conducted two washback studies that focused on the high-stakes English entrance examinations for Japanese universities. He used an experimental design to compare the teaching practices of two teachers in order to see if the entrance examinations pressured teachers to use grammar translation method. Both teachers were giving courses at high schools and also at Japanese cram schools (yobiko). In the high school, he observed them teaching regular and exam preparation classes. At the yobiko, he observed the teachers giving exam preparation

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

117

courses for two different universities. Interviews were held immediately before the observation to gather background information about the teachers. Post-observation interviews were also conducted following each observation. From the data, he found that the washback effect was much weaker than he had hypothesized – washback is only one of several factors influencing teaching practices in class. In fact, Watanabe postulates that the teachers' educational background, their beliefs about effective teaching methods and also, the timing of the observations, that is, how close the examinations were to the time of observation, could be important factors influencing how washback happens.

Ye (1998a, 1998b) presented the results of two surveys. One was administered to 74 EFL teachers from 18 institutions of higher learning, and the other was administered to 174 students at Shanghai Jiaotong University. Based on the results of her questionnaires, she claimed that the CET had not only brought changes to teaching content and teaching methods, but also changed the phenomenon of lecture-based instruction, and increased students' learning initiative and independent thinking. However, her study did not provide sufficient evidence or data to justify her claim. In spite of this claim, she admitted that grammar and vocabulary continued to constitute a considerable portion in CE teaching. It seems that these conclusions are contradictory.

Saif, Shahrzad (1999) carried out a study on theoretical and empirical considerations in investigating washback. This study examined washback as a phenomenon relating to those factors that directly affected the test to those areas most likely to be affected by the test. The goals of the study were: to investigate the existence and nature of the washback phenomenon; to identify the areas directly/indirectly affected by washback; and to examine the role of test context, construct, task, and status in promoting beneficial washback. Theoretically, this study conceptualized washback based on the current theory of validity proposed by Messick (1989, 1996). It was defined as a phenomenon related to the consequential aspect of the test's construct validity and thus achievable, to a large extent, through the test's design and administration.

Given this assumption, a conceptual and methodological framework was proposed that identified 'needs", 'means", and "consequences" as the major focus areas in the study of washback. While the model recognized tests of language

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

118

abilities as instrumental in bringing about washback effects, it highlighted an analysis of the needs and objectives of the learners (and of the educational system) and their relationship with the areas influenced by washback as the starting point for any study of washback. The approach to data collection was both quantitative and qualitative.

The findings of the study indicated that positive washback could in fact occur if test constructs and tasks were informed by the needs of both the learners and the educational context for which they were intended. The extent, directness, and depth of washback, however, were found to vary in different areas likely to be influenced by washback. The areas most influenced by washback were found to be those related to immediate classroom contexts: teachers' choice of materials; teaching activities; learners' strategies; and learning outcomes. The study also revealed that non-test-related forces and factors operative in a given educational system might prevent or delay beneficial washback from happening. Based on the theoretical assumption underlying the definition of washback adopted in this study, many consequences which could not be traced back to the construct of the test were outside the limits of a washback study.

3.2.2 Washback Studies from 2000 to 2005

Cheng and Falvey (2000) conducted a research on "What Works? The Washback Effect of a New Public Examination on Teachers' Perspectives and Behaviours in Classroom Teaching". Hong Kong introduced the Certificate of Education Examination in English to bring about positive washback in classroom teaching. A large-scale research study was carried out over a period of three years to investigate what actually worked with the introduction of the new Certificate of Education in English. The findings of this study indicated that the Hong Kong educational system responded rapidly to the change. Their study found that assessment could leverage educational change and bring positive washback effects to teaching. Washback, as a process, was seen to occur quickly and efficiently in the creation of language teaching materials. Teachers' and students' perceptions of classroom teaching and learning activities were also directly influenced. However, the washback process on the teaching methods that teachers used occurred slowly and reluctantly. The study revealed that the washback effect on classroom teaching

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

119

was limited and superficial. It was postulated that only a combined effort of effective teacher education and materials development could bring about genuine change in classroom teaching. They also put forward some recommendation for promoting washback of the public examination.

Jin (2000) examined the washback effects of the College English Test (CET) Spoken English Test. Questionnaires were distributed to 358 students who took the test in the year of 1999, and to 28 English teachers who worked as interviewers in the test. The questionnaire covered the following areas: students' motivation to take the test, the importance of the test, and its potential washback effects. A large number of students (79.6%) reported that they took the test to have their communicative competence in English evaluated. Most of the students (96.9%) and teachers (100%) thought that it was important to have an oral test in the CET battery. All of the teachers believed that the Spoken English Test would have a huge impact on college English teaching and would promote students' ability to use English communicatively; 92.3% of the students and all the teachers suggested that the test should be accessible to a larger number of students.

The questionnaire also asked the teachers and the students to evaluate the test design, which included test method, test format, test tasks, test time, the reliability of the test, and the rating scale. The results were very positive. The researcher claimed that since the administration of the CET-SET, positive changes took place in college English teaching. For example, many colleges and universities began to pay more attention to improving students' communicative competence; students became more involved in the oral activities in class; and some universities even developed teaching materials that catered to the test. However, there was lack of empirical studies or evidence to support these claims so far.

Chapman & Snyder Jr. (2000) carried out a study on high-stakes testing influences and teachers' classroom methodology. The study in Uganda by Snyder et al. found that changes made to a national examination did not have the desired effect of encouraging teachers to alter their instructional practices, they suggested that it was not the examination itself that influenced teachers behavior, but teachers' beliefs about those changes'

Chen (2002) examined the nature and scope of the impact of the Taiwanese Junior High School English Teachers' Perceptions of the Washback Effect of the

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

120

Basic Competence Test in English. The relational research method was used in this research. The target population was junior high school English teachers. A number of 151 teachers teaching in the 11 -grade were requested to respond to the questionnaire, and focus group interviews. The bivariate correlation and multiple regression analyses were used to analyze the quantitative data. Content analysis using a note-based technique interpreted the qualitative data.

Findings indicated that the public examination associated with educational reform had an influence on teachers' curricular planning and instruction. This washback influence on teachers' teaching attitudes was quite superficial; the washback might influence teachers about what to teach, but not how to teach. It was recommended that longitudinal studies, such as long-term classroom observations, should be conducted in order to explain to what extent washback actually occurs to influence classroom teaching. Findings led to recommendations for teacher professional development, a change of the Taiwanese "academic watch" program, mixed ability grouping, and the addition of oral and aural assessment to the examination. Based upon the findings, this study recommended to: (1) provide teachers with extensive professional development opportunities; (2) change the 'academic watch' policy; (3) practice mix-ability grouping instead of achievement grouping to group students; and (4) integrate assessment into classroom evaluation.

In New Zealand, Read and Hayes (2003) carried out a research to examine IELTS impact. The research was carried out in two phases, moving from a broad overview of the national scene to a specific focus on particular language schools. In the first phase a survey was made of the provision of IELTS preparation in the tertiary/adult sector. They mailed out to 96 language schools throughout New Zealand to collect information on whether schools offered an IELTS preparation course for the Academic Module and, if so, to obtain the basic details of how the course was taught. Of the 78 schools which responded, 77% of them offered IELTS preparation. This compared to 58% that taught English for Academic Purposes (EAP) or English for Further Study (EFS), and just 36% that prepared students for TOEFL.

Their questionnaire was followed up in phases two by 23 interviews with teachers engaged in IELTS preparation at the larger language schools in four of the main centers. The interviews probed the structure and delivery of IELTS preparation in greater depth, as well as exploring the relationship between preparing students for

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

121

the test and preparing them adequately for academic study through the medium of English. The participants reported that students really needed to be at an upper-intermediate level of General English proficiency before being able to benefit from IELTS preparation and have a realistic chance of passing the test, but there was often pressure to accept students whose proficiency was lower than that. Even students who gained the minimum band score for tertiary admission were likely to struggle to meet the demands of English-medium study in a New Zealand university or polytechnic. IELTS courses varied a great deal in the extent to which they could incorporate academic study skills which were not directly assessed in the test. Despite its limitations, the teachers generally recognised that IELTS was the most suitable test available for the purpose.

The study of Hwang (2003) was designed to examine the washback effect of the College Scholastic Ability Test (CSAT), a university entrance exam, on EFL teaching and learning in Korean secondary schools. This study first investigated the relationships among the curriculum, the school textbooks, and the CSAT: (1) the relationship between the curriculum and the textbooks; and (2) the relationship between the curriculum and the CSAT. Second, this study examined if a washback effect from the CSAT existed. This study further discerned the nature of washback and the variable(s) influenced by the washback effect. The results indicated that the curriculum corresponded to the textbooks, while the CSAT did not represent the curriculum, and that there was a negative washback effect of the CSAT on EFL teaching and learning. The variable(s) influenced by the washback effect were negative attitudes that the participants of the study had toward the test.

Hayes, B. M. (2003) investigated the washback effect of the test by studying three IELTS preparation courses offered by language schools at public tertiary institutions in Auckland. The aim of her study was to identify the significant activities in an IELTS preparation class in New Zealand and establish whether there was evidence of washback in the way classes were designed and delivered. Various forms of data-gathering were utilised, including two structured observation instruments, questionnaires and interviews for the teachers, two questionnaires for the students, and pre- and post-testing of the students. In addition, an analysis was made of IELTS preparation textbooks, with particular reference to those which were sources of materials for the three courses. Thus, her study provided a detailed

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

122

account of the range and duration of activities occurring in IELTS preparation courses as well as insight into the teachers` approach to selecting appropriate lesson content and teaching methods.

The findings of her study showed markedly different approaches between the courses, with two focusing almost exclusively on familiarising students with the test and providing them with practice on test tasks. On the other hand, the third course, while including some test practice, took a topic-based approach and differed from the others in the amount of time spent on the types of activities one might expect to find in a communicative classroom. Pre- and post-testing revealed no significant gain in overall IELTS scores during the courses. The study concluded that teachers who designed and delivered IELTS preparation courses were constrained by a combination of factors, of which IELTS itself was but one. Hayes's study highlights the need for further research into appropriate methodologies for washback research, including the refinement and validation of observation instruments, and provides more evidence of the complex impact of tests on both classrooms teaching and learning IELTS.

Linda (2003) carried out a study which aimed at determining the impact of Louisianans School and District Accountability System on students' performance on the state mandated criterion-referenced test. The study was designed to determine the extent to which teachers in the schools in a large urban district in southwest Louisiana turned to instructionally unsound practices in response to a high-stakes accountability system. The specific objectives addressed in this study were to: 1) explore if test scores changed beyond what would be expected given the cohort design of the accountability model; 2) explore if test scores changed teaching methodology; and 3) determine where there had been improved learning and identify those practices teachers used to obtain the positive results. For the qualitative analyses, data was collected from interviews, surveys and observations with 4th grade teachers and principals in the selected school district. Specifically, this study attempted to determine if a measurable increase in student performance on the state-mandated test in grade 4 and determine to what sources the positive change could be attributed. The results of this study indicated that Louisianas accountability system had impacted each school in various ways. There was not only a variation in how these schools perceived accountability, but also a variation in the perceptions of

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

123

teachers and principals with regard to strategies that were being used to prepare students for high stakes testing.

Liu and Dai (2003) conducted a nationwide large-scale study on teacher perceptions of teaching methods, teacher pedagogical knowledge and potential for conducting research, and issues related to instructional innovations and testing. The results revealed that more than 90% of the College English instructors maintained that the CET could not objectively reflect students' communicative competence. They attributed the negligence of aural/oral aspects of language in instruction to the phenomenon of teaching test-related items. They argued that as a test which measured students' linguistic knowledge rather than their abilities in language use, the CET could only encourage students to focus their attention on language knowledge. This, according to them, has led to the test's negative impact. They ended their paper with a call for devising the CET as a criterion-referenced test. They further suggested that subjective questions be increased, and commercialization of the test be avoided. While the data presented in this study might not be taken as evidence of washback, for it was not associated with "the introduction of an innovation intended to cause change" as described by Wall and Horák (2007, p.99), the study provided a window on how Chinese EFL teachers perceived the CET.

Qi (2004) carried out a study by examining the National Matriculation English Test (NMET). In her study, she carried out in-depth interviews and follow-up discussions with eight test constructors, ten senior secondary school teachers, and three English inspectors. Based on the coded data, Qi analysed the structure of the Senior III English course from both the chronological and conceptual perspective using a concept put forward by Woods (1996). For this purpose, data was collected through interview and questionnaire from eight NMET constructors, six English inspectors, 388 teachers and 986 students. She found that de-contextualised linguistic knowledge still had a central place in the Senior II English Course at the expense of communicative meaning and contexts, this despite the decreased weighting on linguistic knowledge in NMET over time.

The findings of her study revealed that the most important reason for the test failing to achieve the intended washback was that its two major functions – the selection function and the function of promoting change – were in many ways in

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

124

conflict with each other making it a powerful trigger for teaching to the test but an ineffective agent for changing teaching and learning in the way intended by its constructors and the authority. Qi's conclusion was that the NMET produced only limited intended washback effects, as teaching of linguistic knowledge was still emphasised and the kind of language use in teaching was restricted to the skills tested in the NMET". Her study also confirmed the circuitous and complicated nature of washback. Finally, Qi suggested that tests might not be a good lever for change – that educational system or school practices would not let themselves be controlled by test constructors. In China, the NMET was not an efficient tool for inducing pedagogical change.

Cheng (2004) investigated the possible washback effects of the Revised Hong Kong Certificate of Education Examination in English (HKCEE) on teachers and students in Hong Kong secondary schools. Her study was a qualitative research conducted through Mixed Method Approach (MMR). She observed 12 high school teachers for 45 lessons. She also conducted a questionnaire survey among 550 teachers and 1700 students. She interviewed an unspecified number of teachers from 1994 to 1995. The ostensible intention of the exam reform was to inspire integrated, task-based teaching. Cheng, however, determined from the questionnaires that although most teachers felt positively about the revised exam that enabled students to use English more practically and effectively, no major changes emerged in terms of actual pedagogic practices, which were still content-based and teacher-centered.

The content of what was taught now focuses more on listening and speaking in accordance with the revised exam. Cheng stated that the change of the HKCEE toward an integrated and task-based approach showed teachers the possibility of something new, but it did not automatically enable teachers to teach something new (p. 164). Cheng's study confirmed Wall and Alderson's (1993) previous findings: while classroom content might change because of a test, the way teachers instructed did not change to any significant degree. The changes noted by Cheng (2005) were "superficial".

Ferman (2004) conducted a study on the washback of an EFL national oral matriculation test to teaching and learning. The EFL test implemented by the Israeli Ministry of Education appears to be one of the studies that included only student data. He found that students' washback behaviours appeared to be influenced by the

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

125

teachers' instructional behaviours with respect to the test. This seemed an important aspect to consider for the Spanish 104 study, as the intent was to observe how teacher behaviours changed over a period of time. Additionally, current educational practice uses student performance as a common way to judge teacher efficacy. If teachers can be judged on how well their students perform on tests, it is relevant to gather data to determine whether or not teacher behaviours related to the tests, in turn influenced student behaviours with respect to the tests.

Han et al. (2004) conducted a survey in China among 1194 English teachers of 40 colleges and universities asking about their attitudes toward the national testing system of the CET at the tertiary level. They found that 37.7% of the teachers thought that the CET pushed colleges and universities to use the passing rate of the test to evaluate their teaching. Over 70% of the teachers did not believe that the test could improve overall English teaching and learning at the tertiary level in China. About 25% of the teachers pointed out that the test encouraged students to guess and to use test-taking strategies, rather than to improve their actual language ability, and 37.8% of the teachers attributed the lack of communicative competence of their students to this test. However, about 70% of the teachers did not want the test to be abolished. From the interviews with some university administrators and English teachers, the researchers found that one reason for this contradiction in attitudes was the time and effort that would have been consumed to design their own test systems and to grade large numbers of test papers. Another concern was the validity issue of a possible self-designed test by an individual university.

In terms of classroom teaching, about 40% of the teachers believed that the CET influenced regular teaching. When asked about a suitable type of a national test for college English teaching, 40% of the teachers thought that it should be a language proficiency test rather than an achievement test, and 45.4% of the teachers suggested that all four skills should be assessed in order to promote students' overall language competence. The teachers were also asked their opinions regarding the relationship between the CET certificate and students' actual language ability. Most of the teachers (77.9%) did not think that these two components were correlated, i.e. having a CET certificate does not necessarily mean that the student has the language competence as required by the College English Syllabus. These findings showed that teachers were doubtful about the validity of the CET.

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

126

Huang, S. (2004) conducted a study on Washback Effects of the Basic Competence English Test (BCET) on EFL Teaching in Junior High School in Taiwan. The data was elicited through questionnaire and interview. The respondents were the English teachers and students. The research questions of the study were on: effects of the BCET on EFL teaching materials, teaching methods, assessment, and students' learning in junior high school. The subjects were 82 English teachers and 351 third-grade students chosen from different junior high schools in central and northern Taiwan.

The quantitative data was analyzed by descriptive statistics to present the mean, standard deviation, and percentage of the responses for each item. Then the interview was transcribed and utilized as complementary opinions. Major findings showed both positive and negative washback effects of the BCET on EFL teaching materials, methods, assessment, and students learning. First, the BCET exerted influence on teachers' decision on selecting textbooks, providing extra reading authentic materials, and adopting realistic audio-visual aids. Huang (2004) believed that the findings of the study might contribute to the improvement of the BCET items. The researcher provided suggestions towards the administration of the BCET and the reformation of the current EFL teaching and learning in junior high school.

Hawkey (2004) conducted a study entitled "A Study of washback regarding the impacts of IELTS, especially on candidates and teachers". In this study, the researcher's main focus was to ensure that the test was as valid, effective and ethical as possible. The instruments were subjected to a range of validating measures including: descriptive analyses (mean, standard deviation, skew, kurtosis, frequency). A total of 572 IELTS candidates from all world regions participated in the study. Findings from the study indicated that 90% of the teachers participating in the study agreed that IELTS influenced the content of their lessons, 63% of the teachers agreed that the examination influenced their methodology. The study concluded that there appeared to be strong IELTS washback on the preparation courses in terms of both content and methodology.

Gu, X (2005) conducted a research in china to explore the relationship between the College English Test (CET) and college English (CE) teaching and learning. The research focused on: the CET participants' perceptions of the test and its washback; the processes of CE classroom teaching and learning, including CET

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

127

washback on CE classroom teaching and learning; and the products of CE teaching and learning. In addition, other major factors exerting influence on CE teaching and learning were analyzed. A number of 4500 CET stakeholders (e.g. administrators, teachers, and students) were involved in the study. Various research methods were employed including classroom observations, questionnaire surveys, interviews, tests and analyses of documents, of 'coaching materials', as well as of CET data and of the examinee output in the CET.

The findings showed both positive and negative washback of the CET. Most of the CET stakeholders thought highly of the test, especially its design, administration, marking and the new measures adopted in recent years. They believed that the positive washback of the test was much greater than the negative washback, and that the negative washback was primarily due to the misuse of the test. However, some CET stakeholders were dissatisfied with the overuse of the multiple-choice (MC) format in the test, the lack of direct score reports to the teachers, the incomplete evaluation of the students' English proficiency without a compulsory spoken English test, and the use of the test as the sole means in evaluating the quality of CE teaching and learning. The study concluded that the issue of the CET washback was complicated and pointed out that the CET was part of a complex set of factors that determined the outcome of CE teaching and learning. The top three factors within the school context were: students' educational background, teacher quality, and administrators' attitudes about the CE courses and the CET.

Lopez, Alexis (2005) carried out a research on the potential washback of the English proficiency test. The study investigated the potential washback of the Integrated Task on classroom practices. The integrated task was a writing task on a new English language proficiency test developed to assess English language learners (ELLS) in grades K-12. This study was conducted in an elementary school in the Midwest. Participants in the study included an ESL teacher, twelve ELLS and thirteen ESL experts. Data was collected using mixed method (MM) approach including a content evaluation of the integrated task, classroom observations, interviews with the teachers and students, think-aloud protocols, and analysis of the students written products. Results of this study highlighted the relationship among the integrated task, ESL writing instruction, and students' writing processes and

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

128

written products. The findings suggested that there were matches and mismatches between the task and classroom practices. Lopez (20050 commented that this alignment could potentially inform test developers about changes that could be made to the task.

HUANG, L. (2005) carried out a research on the nature of the washback effects of the Senior Secondary School Entrance Examination (SSSEE) English oral test. The study showed that washback is a complex phenomenon and it could be conceptualized via a multidimensional model. His study presented preliminary research findings related to the washback effects of the oral test on teaching. The data was collected via focus groups and questionnaires. 51.1% of the teachers said that they often provided students with lessons specifically focusing on speaking skill development. 82.8% indicated that the administration of the oral test had raised their awareness of teaching communicatively. 42.2% of the teachers reported that they often used specially tailored materials for the oral test. 48.6% declared that it was necessary to use special coaching materials for the oral test. 18.4% did not think it was necessary.

This suggested that washback of the oral test on the development and introduction of new teaching materials did exist but was not strong. This result suggested that the washback on teachers' perceptions of the importance of speaking teaching was strong, which was consistent with the findings from the focus group. As for the teaching methods, 39.7% of the teachers reported that they used computer software to help speaking training. It was found that the washback of the oral test on teaching existed. In the questionnaire, 60% of the teachers indicated it had a substantial impact on their teaching. In the focus group, similarly, some teachers acknowledged that the oral test changed their teaching routines and methods. In addition, they employed the test type of the SSSEE oral test in their daily teaching.

Caine, A. (2005) carried out a study to examine the effects of existing EFL examination on teaching and learning in Japan. An attempt was made to determine the extent and nature of washback resulting from this new speaking test. The subjects consisted of teachers and learners taken from the upper secondary education (i.e. high school) sector in Japan. Most of the research was conducted at one participating school – a private high school of 486 students in the south of Japan. In addition to classroom observation, teacher and student questionnaire surveys were

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

129

also administered in order to measure the washback effect of EFL tests currently taken in the sample context. The data was collected from the teacher questionnaires and the classroom observation. However, additional data was also collected from teachers working at public high schools in the area. The study focused on the 'mismatch' that occurred between the levels of curriculum planning and actual classroom implementation.

The results of this study suggested that it was possible to improve learning by employing direct testing techniques. It was proposed that future research should be conducted using a large sample group and that data should be collected longitudinally. The study commented that more direct testing techniques were needed to be employed in a larger number of "high stakes" examinations to effect the changes on teachers and teaching English as a foreign language

Manjarres (2005) conducted a study that intended to test washback within a high-stake test. The general objective of the study was to describe the washback effect of the English national examination held at public schools in Colombia. The central question of the study was whether the English Test had any washback effect on teaching English, and whether the exam tested students' grammatical and linguistic competence. The researchers analysed the tests students took in 2003 and 2004. The gathered data was then compared with the classroom practices recorded from the observations, (five lessons were observed), an interview with three students, a formal interview with an English language teacher, and an interview with the latter together with another English language teacher of the school. Manjarres (ibid) advocated that the central question of this study was whether the English Test had any washback effect on the teaching of English in the specific context of this study, which could be considered a representative case of public schools in big towns in the northern part of Colombia.

The results of the study showed a positive relationship between the exam and the teachers, that was, English language teachers adjusted their strategies in order to meet students' expectations, this was also noticeable when teachers depended on other materials to perform better in the classroom (i.e. previous test formats). The study also showed that teachers were not familiar with how to develop students' communicative competence. The study found that listening and speaking skills were

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

130

not evaluated in the exam. In addition, teachers' main focus was on developing students' grammatical skills.

Ying, Y. (2005) investigated the washback effects of the Spoken English Test (SET). The findings of Ying's study showed that teachers used different approaches and methods when teaching, and that they looked at the examination as crucial and important. However, SET teachers seemed to concentrate on communicative competence, they neglected the usage of grammar and translation. The SET examination was set to measure students speaking skills. Textbook evaluation revealed that the influence of SET on the design of the textbook series only occurred at the superficial level, i.e., it influenced the contents and formats of the speaking elements in the textbook series. This indicated that the design of the textbook series received more influence from the teaching syllabus than from SET, which was confirmed by the interview with the textbook writer. The findings of the study brought insights to the washback effect of tests on teachers, in terms of changing their teaching methods when teaching for a high-stakes exam. The findings might also stimulate textbook writers to pay attention to the overall construct of grammatical exercises in the development of English textbooks.

3.2.3 Washback Studies from 2006 to Date

Green, A. (2006) conducted a study to investigate washback to outcomes by comparing learner performance on the three course types. IELTS writing tests were administered at course entry and exit and a gain score for each learner calculated as the simple difference between these entry and exit scores. As both participant and process variables other than course type might account for any differences in mean score gains found in the study, data relating to course length, course intensity (hours of study per week) and individual characteristics, beliefs and attitudes considered likely to mediate washback were accessed through questionnaires and course documentation.

The participants of the study were international students preparing for academic study at fifteen institutions in the UK. These institutions were selected following an earlier survey of UK course providers. They were willing to participate and were conveniently located. A number of 663 students participated in the

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

131

research. A total of 476 (71.8%) students completed both entry and exit forms of the IELTS academic writing test. Paired t-tests were used to investigate whether learners had made score gains following their courses. The results indicated that a significant gain in writing scores had indeed occurred on all three course types Taken as a whole, the learners improved their IELTS academic writing scores by an average of 0.207 of a band on the nine-band IELTS scale. This indicated that students with higher initial writing scores made less gain than their lower scoring counterparts. Other features that displayed relatively high correlations with writing score gain included the grammar and vocabulary measures, use of test-taking strategies, self-assessed improvement in writing ability and self-confidence in English writing ability.

Saif (2006) carried out a research to explore the possibility of creating positive washback by focusing on factors in the background of the test development process and anticipating the conditions most likely to lead to positive washback. The study focused on the washback effects of a needs-based test of spoken language proficiency on the content, teaching, classroom activities and learning outcomes of the ITA (international teaching assistants) training program linked to it. As such, the conceptual framework underlying the study differs from previous models in that it includes the processes before test development and test design as two main components of washback investigation. The analysis of the data – collected from different stakeholders through interviews, observations and test administration at different intervals before, during and after the training program – suggests a positive relationship between the test and the immediate teaching and learning outcomes. The results obtained from interviews, observations, and quantitative analysis of test scores suggested that the ITA test had some influence on classroom-related areas such as teaching content, teaching methodology, and students' learning. The results also revealed that the depth, extent and direction of the effect differed with the affected area. The content of teaching seemed to be the area showing changes directly triggered by the test. This was in line with the results of previous studies on washback (see, for example, Wall and Alderson, 1993; Alderson and Hamp-Lyons, 1996; Cheng, 1997) that found the content of language teaching as the area readily susceptible to change as a result of tests. Class observations and teacher interview revealed that the teacher's adaptation for the ITA course of the materials available to her was based on two factors: the objectives of the course and her impression of the

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

132

ITAs' language abilities after the first administration of the test. There is, however, no evidence linking the test to the policy or educational changes at an institutional level.

Green, A, (2007) investigated whether test preparation classes were advantageous in assisting students trying to improve their IELTS writing scores. There were three sub-groups: 85 participants attending IELTS preparation courses, 331 in the pre- EAP course, and 60 in combination courses. All participants were asked to take the IELTS grammar/vocabulary tests at the beginning and end of their 4-to-14-week courses. Questionnaires examining participant and process variables such as learner background, motivation, class activities, and learning strategy use were completed after the pre and post tests. Inferential statistics were adopted and revealed that no clear advantage for focused test preparation. In addition, score gains were found primarily among two groups of learners: those who planned to take the test again, and those who had low initial writing test scores. Washback to the learner rather than washback programme had more to do with the improvement in students' test scores. These findings had two implications: first, test-driven instruction did not necessarily raise students' scores. A more beneficial way to improve students' scores might be to integrate material covered on the test with regular teaching. Second, concerning this point, intentions for taking the test needed to be clear to both students and teachers to foster English learning.

Wang, H. (2006) conducted a study on an implementation study of the English as a foreign language curriculum policy in the Chinese tertiary context. This study explores the implementation of the mandatory national college English curriculum within a Chinese tertiary context. Using a mixed methods approach, she conducted the study by engaging three groups of participants. She interviewed four national policymakers in terms of syllabi, textbooks, and tests to identify the intended curriculum. She interviewed six departmental administrators to determine their perceptions of the national language policies and their roles in ensuring the implementation of these policies. She conducted surveys to discover 248 teachers' perceptions of the intended curriculum and uncovered the factors affecting their implementation activities in the classroom. By observing two teachers' classrooms and through follow-up interviews, she also examined how the language policies were being interpreted at the grass-roots level.

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

133

The findings revealed a discrepancy between policymakers and administrators and between policymakers' intentions and teachers' implementation. Policymakers designed general, open-ended, and abstract policies to offer local universities and teachers some flexibility and autonomy when they put those policies into practice. However, administrators as intermediary individuals between policymakers and implementers apparently interpreted the open-endedness of the curriculum policies differently than the policymakers had intended. Instead of using the built-in flexibility to tailor methods of helping students gain proficiency, they placed their emphasis on only one outcome—students' good scores on the national English test. They also failed to support their teachers in understanding the policies by not providing necessary resources to help them implement the policies fully.

Furthermore, the research uncovered five external and internal factors as significant predictors of teachers' implementation: resource support, teaching methods (communicative language teaching and grammar-translation method), teaching experience, language proficiency, and professional development needs. Classroom observations and interviews revealed that teachers failed to implement what was expected from policymakers in the classroom. Rather, they conducted teaching based on the classroom and political reality. Their factors were mainly student factors and the departmental factor. The implications of this study were pointed to the importance of the intermediaries, the department heads, in both providing the necessary pressure (motivation) and support (resources) necessary for the implementation to take place.

Shih, C. (2007) investigated stakeholders' perceptions of the Taiwanese General English Proficiency Test (GEPT) as well as its washback on schools' policies, teaching, and English learning. The research sites were the applied foreign language department of university of technology (School A) and an institute of technology (School B). The latter school required day-division students to pass the first stage of the GEPT intermediate level or the school-administered make-up examination, whereas the former did not prescribe any GEPT requirement. In each department, he reviewed its records and interviewed the department chair, 2 to 3 teachers, 14 to 15 students, and 3 parents or spouse of 3 participating students. Shih also observed one of the courses taught by each interviewed teacher as well as the self-study center 2 hours weekly for 8 selected weeks out of one semester. One

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

134

exception was a GEPT Preparation course at School B, which he observed for a whole semester.

The findings of the study indicated that the GEPT had a little or no impact on teaching at both schools, except for courses at School B which were germane to the school's GEPT policy. Although the GEPT generated various degrees of washback on English learning at both schools, there was an absence of long-term systematic preparation for the test. A handful of students prepared for the GEPT two months before the test, whereas some students had no preparation whatsoever. Some teachers believed that the GEPT was valid and reliable, whereas others had neutral or negative perspectives on these issues. Participating students believed that the GEPT had gained public credibility. However, they still pointed out several issues and problems with the test. The results of the study indicated that the existing theories or models did not fully explain the washback of tests on learning. He therefore proposed a new, tentative washback model of students' learning to delineate this subject. Moreover, although results seemed to discourage using the GEPT as a degree requirement or other gate-keeping purposes, he suggested several guidelines for those schools, which, out of some considerations, should adopt the GEPT for these high-stakes purposes.

Shih, C. (2008) conducted another study to compare one private technical college in Taiwan that required English majors to pass the elementary level of the General English Proficiency Test (GEPT) with a similar private technical college, which had no such graduation requirement. The GEPT was commissioned by Taiwan's Ministry of Education in 1999 and is a criterion-referenced test that reputedly measures writing, speaking and listening skills. Interviews with 2 department heads, 6 teachers, 30 students, and 3 family members were conducted. Observations were made for a semester in test-preparation classes or in classes that taught skills tested on the GEPT. Departments' policies regarding the GEPT exit requirements were also reviewed. The findings indicated that the GEPT had elicited a varying but minor impact on learners at both schools, although a slightly higher degree of washback was found at the school with exit requirements. In addition, Shih generated a new washback model of students' learning. This model includes extrinsic, intrinsic, and test factors to help depict the complexity of learning washback.

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

135

Shih, C. (2009) investigates that how test change teaching. The purpose of this study was to investigate the washback effects of the General English Proficiency Test (GEPT) on English teaching in two applied foreign language departments in Taiwan. One had prescribed its GEPT requirement to its day-division students whereas the other had not. Overall, the GEPT did not induce a high level of washback on teaching in either department. Only courses which were linked to the departmental GEPT policy and whose objectives were to prepare students for the test were significantly affected. The results of his 16 hours of observation showed that the GEPT had an impact on Don's teaching content as well as mid-term and final examinations, but not on other aspects of his teaching. His teaching material was a monthly GEPT magazine that was available in local bookstores. Mid-term and final examinations were simulated GEPT examinations, which were produced by the same GEPT magazine publisher. On the other hand, Don never mentioned the GEPT explicitly in class, never offered GEPT relevant information to his students, and did not instruct students in any test-taking strategies.

The results of my observations were mostly congruent with Don's testimony in his interview. He also believed that his teaching content was relevant to the GEPT, and the mid-term and final examinations were mock GEPT tests. However, he rarely coached students in test-taking skills and seldom offered students GEPT-relevant information. The findings suggested that micro-level contextual factors (for example, the objectives of the course) and teacher factors had a greater impact on teachers' instruction. Finally, on the basis of current understandings of washback, Shih proposed a new tentative model to portray the washback of tests on teaching.

Karabulut (2007) carried out a study on "Micro level impacts of foreign language test (university entrance examination)" in Turkey. The purpose of this study was to find out whether the foreign language examination---university entrance test---influenced the way teachers taught and students learnt in senior three classrooms (the last grade of high school) in Turkey. Secondary goal was to see the outcomes of teaching to the test and attitudes of different stakeholders towards the test and senior three English teaching in general. For this study, data was collected through online surveys; and participants comprised four major groups. Senior three high school students and English teachers were invited to participate to find out the nature and the scope of washback, while college students and professors were asked

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

136

to participate to investigate the outcomes of teaching to the test. Descriptive statistics were used to analyze the responses of the participants. The results suggest that the test is a major factor determining the flow of English lessons in senior three classrooms. The classroom materials that were reported by both students and teachers including mock tests, commercial exam preparation materials and sample test questions directly served to the purpose of practicing for the test and indicate the relative effect of the test on language learning.

The results suggested that high school students and teachers focused more on the immediate goal of language learning which was to score high on the test and be admitted to the university by cramming for the test, and learning and practicing the language areas and skills that were measured on the test (grammar, reading, and vocabulary items) and ignored the ones that were not tested (listening, speaking, writing). The teachers and college students, on the other hand, felt the enough practice especially in productive skills should have taken place in the classroom. The respondents opined that long- term goal of language learning should be to improve the ability to use the language. Based on the gap reported by these different stakeholders, findings led to recommendations for a change in the curriculum and in the format of the test towards a more communicative and integrative one.

Huang, C. (2007) carried out a study to explore the washback effects of the General English Proficiency Test (GEPT) on English-language teaching and learning in an EFL context. Moreover, it aimed to investigate how the GEPT influenced current English-language teaching and learning. The data was collected through a questionnaire and interview from English teachers and students. In the study, convenience sampling was adopted and the participants were nine English teachers and 306 students chosen from nine junior high school classes in northern, central, and southern Taiwan. Both quantitative and qualitative data was used for the study. The quantitative data was analyzed by descriptive statistics to present the mean, standard deviation, and percentage of the responses for each item. On the other hand, the qualitative data collected from the interview was transcribed and utilized as complementary opinions.

The results of the study revealed that both positive and negative washback effects of the GEPT on EFL teaching and learning occurred. The students admitted that when examination got closer they studied harder than before. They were

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

137

motivated to learn English and thus became autonomous learners. Moreover, students were aware of the significance of fostering the four language skills. Further, getting the GEPT certificate gave students a sense of achievement and gave them a competitive advantage when applying for senior high schools or finding a good job.

Mohammadi, M. (2007) carried out study on the washback of the High-Stakes Testing on teaching. This research aimed at conducting a survey of the washback effect of MA Entrance Examination on teachers' methodology and attitudes. 45 subjects, all of whom university professors, were selected using convenience random sampling. Then, a validated researcher-made questionnaire was administered. To have more reliable data, some were randomly selected for interview so as to cross-check the data collected through questionnaire. The data analysis revealed that the majority of the subjects were positively influenced by the examination. Moreover, they were fully aware that their methodology and attitudes were gradually set to the demands of the examination.

Retorta, S. (2007) carried out a study entitled "The washback effect of the Federal University entrance examination of Panama the teaching of the English language in secondary schools of Panama: an investigation of public and private schools as well as cramming courses". The objective of the study was to investigate whether the English test of the University Entrance Examination of UFPR set off the washback effect in the teaching/learning of the language in public and private high schools as well as cramming courses and, if so, what effects were they. In order to meet these objectives a qualitative investigation was conducted in which various voices of the school community were heard such as the participants of public schools (urban and rural), the private schools and the cramming courses (private and free ones). Since there was an intention of having a multi-perspective of the phenomenon, the scenarios were chosen because of the great social inequalities of that country and, therefore, stakeholders were also selected for interviewed. The data was triangulated, analysed and discussed in descriptive and statistical ways. Retorta (2007) also conducted interview and class observations for collecting qualitative data for the study.

The results of this study showed that there was no washback effect of the English test of the University Entrance Examination of UFPR in public schools. What helped set the teaching goals of the discipline were the contents suggested in

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

138

the didactic books adopted in each school. In the other scenarios, the washback effect was observed. The positive effects were the motivation of the directors and teachers to search for information about the test; motivation of the students to study harder to pass the test; the test was used to set clear teaching objectives and reading began to be taught. The negative effects were: anxiety of the participants of some scenarios and curriculum narrowing. This study offered a theoretical contribution when it helped understand a bit more about the washback effect; methodological contribution due to the research design which was innovative and broad and, finally, the study intended to offer a set of information which can give support to the teaching and evaluation of the English discipline in high schools in Panama.

Tsagari (2007) conducted study entitled "Investigating the Washback Effect of a High-Stakes EFL Exam in the Greek context: Participants' Perceptions, Material Design and Classroom Applications". This research project was an attempt to examine the washback effect of a high-stakes examination on the teaching and learning process that took place in the intermediate level classes leading to that level. The researcher interviewed 15 native and non-native EFL teachers, actively involved in teaching FCE. The results led to detailed analysis of textbook materials using a specially-designed instrument. The analysis of the data showed that the exam did influence the materials teachers used when teaching, but it did not show any washback effects upon teachers teaching methods. Implications from Tsagari's study showed that other factors beyond the exam, such as the exam designers understanding of the underlying principles of the exam and their ability to create an affective exam through the materials used, seemed to play a greater role in determining the influence of the exam rather than the exam itself. The final part of the study looked at the effects of the exam reported by students. The analysis of the data showed that students' attitudes and feelings as well as their motivational orientations towards learning the language were affected by the examination.

Tsagari (2009) conducted study which was carried among the 54 EFL teachers and 98 EFL students of various ages and levels of proficiency at two different private language schools in Athens. The results showed that teachers and students did indeed think that language testing had an impact on teaching and learning, although they were not all in agreement as to what that impact was. Several things stood out from the results of the teachers' surveys. First of all, teachers were

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

139

divided in their agreement with the statement "exams help improve classroom teaching" (37% agree, 49% disagree and the remaining 14% don't know). Interestingly, more than half of the teachers (57%) replied that they did not think exams related well to communicative language teaching. Also, 55% of teachers agreed that exams helped give students confidence, but they also overwhelmingly agreed (92%) that exams also caused students anxiety.

The students' surveys revealed, not surprisingly, the majority of students agreed that exams were very important and useful to them (89%), that exams had a positive effect on teaching (66%), on learning (69%), on materials (69%), and the perceived attitude of the teacher (62%). They were less in agreement on the impact of tests on learner attitudes (44% reported a positive or strong positive impact, 20% didn't know, 26% reported some negative impact, and 10% a strong negative impact). The majority of students (70%) unfortunately agreed that exams do cause them anxiety. The mixed results of this survey showed that washback was a "complicated equation" involving teachers, students, materials, attitudes and perceptions.

Choi, I. (2008), in his study, provided an overview of the impact of standardized EFL tests on EFL education in Korea. The study presented the status quo of EFL testing in the Korean context; explores the nature of the EFL tests prevalent in the EFL testing market; and investigates the overwhelming washback effects of EFL tests on EFL teaching based on a survey of stakeholder viewpoints. The overall findings of the survey revealed that the majority of stakeholders (i.e. test-takers and teachers) did not think favorably of the EFL tests due to negative washback effects on their EFL learning and teaching. The survey also showed that considerable numbers of young students were under unwarranted pressure to take the EFL tests and that secondary education put too much emphasis on preparation for the college entrance exam. Most respondents had negative views of the tests in terms of the mismatch between test scores and English proficiency and the failure of multiple-choice EFL test preparation to induce productive English skills. Some respondents voiced complaints about the financial burden caused by mandatory submission of test scores for graduation and employment.

The study of Wall and Horak (2008) focused on the role of communication in creating positive washback. Their study was designed to find out what

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

140

examination designers would say about the role of communication in their efforts to promote positive washback; to find out what teachers would say about the success (or otherwise) of examination designers in communicating what they desired. Data was collected through the online questionnaire. The study found that 82% of test designers discussed washback; 78% of the respondents documented their intentions; 47% of teachers didn't know if the exam was meant to encourage washback. The researchers found that teachers usually did not understand the nature of tests and encouraged testers to communicate their intentions so that teachers and learners could prepare for new kinds of assessment.

Al-Jamal and Ghadi (2008) examined the nature and scope of the impact of the English General Secondary Certificate Examination (GSCE) on secondary language teachers in Al-Karak district located in Jordan. The purpose of this study was to investigate how English language teachers in Al-Karak district who taught second secondary students perceived the impact of the GSCE on their selection of teaching methods. The target population was English language teachers teaching the second secondary class in Al-Karak District in the scholastic year 2006/2007.

A survey questionnaire, which consisted of Likert- Scale items, was used in order to collect the required data. The questionnaire was divided into two parts. The first part of the study aimed at measuring how the GSCE affected English language teachers' method selection in terms of four domains: activity/time arrangement, teaching methods, materials teachers would use in the classroom and content teachers would teach. The second part of the questionnaire, however, investigated the effect of other factors related to the GSCE on teachers' method selection in terms of four domains: students' learning attitudes, teachers' professionalism in teaching, teachers' perceived external pressure in teaching, and perceived importance of the GSCE.

Findings of the study indicated that both the GSCE and the other related factors have affected English language teachers' method selection with a slight statistical difference in favour of the GSCE washback effect. Another indication obtained from the study was that English language teachers in Jordan used the grammar-translation method in teaching English. The results also showed that two types of washback existed in secondary schools in Al-Karak namely: positive and negative washback. In light of the results, the present study recommended that: (a)

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

141

teachers' should be provided with professional development opportunities; (b) teachers' monitoring and evaluation policy should be reconsidered; and (c) GSCE should integrate oral language skills as well. They concluded their study with some recommendations for promoting positive washback.

Jou, C. (2008) accomplished a study on the perceptions of the test of English in a private university in northern Taiwan.. The purpose of this study was to investigate the perceptions of the Test of English for International Communication (TOEIC) and its impact on the school's policies, teachers' teaching, and students' English learning study. The researcher applied both qualitative and quantitative methods for data collection. Interviews were conducted with the chairperson, three teachers, and 8 students of the Department of Applied Foreign Languages of the university. A questionnaire was administered among the respondents of the Department. Besides, formal records, meeting minutes, and official documents concerning about TOEIC were also assembled for analysis. The study lasted for around a year. Results were categorized, transcribed, calculated, analyzed, discussed and described in statistic figures.

The major findings of the study revealed that TOEIC's impact was enormous and decisive. First, it affected the school authorities to make the policy of adopting a TOEIC 650 score as a threshold for the English majors in the Department of Applied Foreign Languages. The enactment of the TOEIC 650 policy brought about a series of measures and actions which in turn had directly or indirectly affected the teaching and learning in the Department. Second, the TOEIC washback on teaching at the Department ranged widely from a high degree of impact to no impact at all, depending chiefly on whether the course was directly related to TOEIC. One teacher has even created a TOEIC vocabulary learning system and put it up online for the students to use free. It was found that TOEIC generated different degrees of washback on individual students' learning in the Department. It had little or no impact on some students, but motivated a few others to study English for at least a period of two or three months. It was also found that some of the students did not seem to have been affected at all by TOEIC and the related TOEIC activities held by the Department.

It was also found that quite a high percentage of the students want to take TOEIC in their college years because was a threshold and they believed that a

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

142

certified high TOEIC score was helpful to their job seeking and further advanced studies after graduation. Pedagogical implications and suggestions were put forwarded for the policy makers, teachers and students on the one hand, and to educational administrators, teachers and educators in Taiwan on the other hand.

Mousavi and Amiri (2009) conducted a study on the washback Effect of TEFL University Entrance Exam on Academic Behavior of Students and Professors. The study was an attempt to investigate the washback effect of the Knowledge test of TEFL MA University Entrance Exam on students and professors. This section of TEFL MA UEE consists of three parts. They are related to the three areas of Linguistics, Testing, and Methodology. To this end, an observation checklist and two questionnaires, one for professors and the other one for the students based on the underlying theories of washback were developed. A total of 32 professors, 210 students, and 13 Linguistics answered the questionnaires. Testing and Methodology classes were observed. Finally, to find the answers to research questions, the Chi – square test and frequency analysis were performed through SPSS. The result indicated that TEFl MA UEE had negative washback on students and professors academic behavior.

Latimer, G. D. (2009) conducted a research on "Washback effects of the Cambridge preliminary English test at an Argentinean bilingual school" in Argentina. This study documented the overall English language program at one Argentinean bilingual school and examines, in particular, the effects the Cambridge ESOL exams upon its curricula, its teachers and upon language learning. This ethnographic research included broad-based observations, conducted over three years, and a five-month investigation of the Cambridge Exams' impact on teaching and learning at this bilingual school. The research found both positive and negative washback effects on language learning. In short, the Exam works against the language development the institution aspires to foster.

Mizutani, Satomi (2009) investigated the mechanism of the phenomenon known as washback in the context of a new national standards-based assessment system in New Zealand, particularly focusing on the area of the teaching and learning of Japanese as a foreign language. The National Certificate of Educational Achievement (NCEA) was progressively implemented across all subjects in the final

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

143

three years of secondary schooling from 2002. It replaced norm-referenced assessments and aimed to function as assessment for learning as well as of learning.

The research consisted of three studies. Studies One and Two investigated washback effects of NCEA as perceived by teachers and students of Japanese, and beliefs about NCEA which contributed to the washback effects. This large-scale study involved teachers and students of Japanese, French, History, and Mathematics. Teacher and Student Questionnaires were developed to investigate washback of NCEA and beliefs about NCEA, Teaching, Learning, and Teacher Efficacy as well as to collect relevant background information on the participants. The study revealed that some contextual factors played a role in mediating certain types of beliefs and washback effects. The results also confirmed that positive washback was promoted when participants' beliefs were in line with the intentions of the assessment. It is concluded that, for educational reform through assessment change to be successful, stakeholders' beliefs about the role of assessment might need to be altered. A model was presented to describe the mechanism of washback, showing how washback could be mediated directly and indirectly by contextual factors and beliefs.

Li, Hongli (2009) carried out a study entitled "Are teachers teaching to the test?: A case study of the College English Test (CET) in China". This study aimed at finding out whether teachers were truly teaching to the test and the potential reasons involved. In order to gain deeper and more focused insight into the influence of the CET on classroom teaching, only its writing section was examined. Based on data collected from some students and teachers at a University in Beijing, China, it was found that the overall influence of the CET writing was not as substantial as what was claimed. Due to different stakeholders' perceptions of the CET, the influence on teachers was weak and indirect compared to a stronger and more direct influence on students. Also, teachers did not teach to the test due to the lower priority of writing among the four skills of language. The relatively low requirement of the CET writing and its restrictive testing format also prevented the teachers from teaching to the test. It was found that the teachers' lack of professional training and some logistic factors outweighed the influence of the CET writing. It was pointed out that teacher factors might outweigh the influence of the CET. Thus, the researcher recommended that teacher should be provided training to improve the efficiency in classroom teaching.

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

144

Turner (2009) carried out a study entitled "Examining washback in second language education contexts: A high stakes provincial examination and the teacher factor in classroom practice in Quebec secondary schools". The participants were the ESL secondary teachers in the French school system in the province of Quebec in Canada. The main research question was: How do teachers mediate between classroom assessment activity and preparing students for upcoming external exams? The findings of the study indicated that teachers used common overall approaches to teaching, but there was variation in individual practice. When first introduced to the new exam material, teachers used a formative assessment approach. As the exam time neared, their practice evolved into a summative assessment approach. This phenomenon demonstrated an interfacing or 'blurring' of formative and summative assessment in an attempt to align classroom and external exam assessment. Implications were discussed pertaining to a coherent education system across curriculum, teaching, learning and assessment. He suggested further intensive study on the areas.

Silva, de Oliveira (2009) carried out a research on "Washback effect of achievement testing in Brazilian regular education: keeping an eye on motivation to learn EFL" in Brazil. The aim of this research was to approach the interrelation between assessment and motivation to learn EFL. The specific aims were to know: what was the effect of formative assessment on the students' motivational orientation to learn EFL, and how formative assessment affected the students' awareness of learning and competence. In order to achieve these aims, ethnographic research methods were employed to describe students' perceptions and motivational orientations facing assessment, which was essentially summative at first, then combined with formative assessment, introduced in the second quarter.

The findings of the study revealed the complex relation between assessment and motivation, which was mediated by the teacher. Results also showed that high achievers changed little in their perceptions and motivation to learn EFL, being intrinsically motivated throughout the school year. Medium achievement students showed changes in their perceptions and motivation, revealing flexible motivational orientations. Finally, low achievers showed small changes in their perception and motivational orientations, which are certainly meaningful considering their low levels of motivation at the beginning of this research. The study concluded with

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

145

some theoretical and practical implications for the studies in assessment and in motivation and for the teaching and learning scenario of English as a foreign language in Brazilian regular schools.

Hsu, Hui-Fen (2009) conducted a study on the impact of implementing English proficiency tests as a graduation requirement at Taiwanese universities of technology. The research sites were non-English departments of Taiwanese universities of technology, which were divided into two groups. One of the groups (Group 1) required non-English major students to pass one of a set of English proficiency tests at a specified level as a graduation requirement, whereas the other group (Group 2) did not prescribe any English graduation requirement. In each group, 27 to 28 teachers and 300 to 321 students completed questionnaires. Two teachers from each group, along with three departmental directors and three advisory committee members within the Taiwanese Ministry of Education, were interviewed. Two lessons taught by each interviewed teacher were also observed.

Findings of the study indicated that the policy of implementing English proficiency tests as a graduation requirement had a superficial or at times no impact on teaching for both groups, with a slightly greater impact on Group 1, who complied with their university's policy of English graduation requirement. Although the majority of Group 1 teachers, departmental directors and advisory committee members had generally positive attitudes towards the policy, teachers' fundamental beliefs about English language teaching and learning were not changed. The new policy influenced what the teachers taught, but not how they taught. In addition, the teachers, departmental directors and advisory committee members pointed out several issues and problems with the diffusion and implementation of the educational innovation.

The researcher found that the teachers and educational administrators nevertheless were aware of the problems they currently faced and appeared determined to resolve them. The results seemed to argue against using English proficiency tests as a degree requirement or for other gate-keeping purposes. Guidelines were also proposed for those universities which wanted to adopt the English proficiency tests for these high-stakes purposes.

Wang, J. (2010) accomplished a research to explore the washback effects of the CET (College English Test) on teacher beliefs, interpretations and practices, and

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

146

in particular seeks to discover the way the 'teacher factor' was manifested in the washback phenomenon. It also investigated the pedagogical as well as the social and personal complexities influencing teachers' beliefs and interpretations and practices. This study answered the research question: What role does the 'teacher factor' play in washback in the Chinese university context? Participants were 195 tertiary-level EFL teachers of the non-English programs.

The main purpose of this study was to investigate whether tests constitute a major constraint on CE (College English) instructional innovation in China. In addition, the intent of the study was to find out what aspects pertinent to this factor (e.g., teacher beliefs, teacher knowledge, experiences) present the major barrier to the implementation of instructional change. A mixed methods approach combining both qualitative and quantitative methods of data collection and data analysis was adopted in this study. A teacher survey and in-depth case studies (through focused group/individual interviews and classroom observations) were used to collect data. Data was analyzed in two phases. Qualitative analysis involved the use of constant comparative method, while quantitative analysis in this study involved descriptive statistics and inferential statistics.

The findings of the study suggested that the CET coupled with various interrelated components of the 'teacher factor' is involved in fostering the washback effect. Given the complexities underlying the washback phenomenon, the educational change carried out in curriculum and assessment was not sufficient on its own to entail teacher change in terms of pedagogical strategies. It appeared that for fundamental changes in teacher practice to occur, they must be accompanied by other changes in teachers' knowledge, beliefs, attitudes and thinking that inform such practice. It was hoped that the issues identified in this study would serve to inform educational authorities, test designers and teachers, and serve as an impetus to upgrade EFL teaching in China.

Muñoz and Álvarez (2010) conducted a research to determine the washback effect of an oral assessment test on some areas of the teaching and learning of English as a Foreign Language (EFL). The research combined quantitative and qualitative research methods within a comparative study between an experimental group and a comparison group. Fourteen EFL teachers and 110 college students participated in the study. Data was collected through the teacher and student

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

147

surveys, class observations, and external evaluations of students' oral performance. The data was analysed using descriptive statistics for qualitative information and inferential statistics to compare the mean scores of the two groups by One Way ANOVA. Results showed positive washback in some of the areas examined. The implications for the classroom were that constant guidance and support over time were essential in order to help teachers use the system appropriately to create positive washback on teaching and learning.

Jin, Y. (2010) conducted a research in china to investigate language testing with a reference to the English teachers. This study was designed to investigate the training of tertiary level foreign language teachers in China with a focus on language testing and assessment courses. A nationwide survey was conducted among 86 instructors of such courses for an overview of the current situation in terms of the instructors, teaching content, teaching methodology, student perceptions of the courses, and teaching materials.

The findings of the study revealed that the courses adequately covered essential aspects of theory and practice of language testing. However, educational and psychological measurement and student classroom practice received significantly less attention. Comparison of the teaching content of the different types of courses did not show major differences. Yin Jin (2010) put forwarded some suggestions to highlight some under-addressed aspects of the teaching content and to set up a network of teacher-testers to create opportunities for practitioners to exchange experiences, professional knowledge and skills.

Barnes, M. M. (2010) investigated washback of a high-stakes English language proficiency test, the Test of English as a Foreign Language Internet-Based Test (TOEFL iBT), on general English and TOEFL iBT preparation courses in Vietnam. For the study, the researcher observed and interviewed four teachers. Teaching materials were also collected from four educational institutions in Vietnam. The study revealed that the TOEFL iBT influenced both what and how the teachers taught, particularly in TOEFL iBT preparation courses. Barnes (2010) believed that the findings of this study had important implications for teaching and learning in Vietnam.

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

148

3.3 Conclusion

The research evidence discussed above illustrates that washback is a highly intricate rather than a simple and a monolithic phenomenon. Over the past decade, there has been a considerable amount of research on washback. This domain of research seeks to answer, in one form or another, one fundamental question – how testing influences teaching and learning. All the research studies reviewed above have provided us with a steady accumulation of knowledge about the nature of washback. However, despite numerous positive qualities demonstrated in the above-mentioned washback studies, we have noticed that they are limited to some extent.

The findings of the washback research discussed above have been inconclusive. Some studies found that teaching content was more likely than teaching methodology to be influenced by tests (Cheng, 1999). Others found that tests influenced both teaching content and teaching methodology, but the extent of the influence of the tests varied from teacher to teacher (Alderson & Hamp-Lyons, 1996; Watanabe, 1996) as well as from student to student (Andrews et al., 2002). The unpredictability of washback effects led researchers to assume that these findings may be due to the variability of the educational contexts of teachers and students. The argument in this thesis is that these washback effects may be powerfully mediated by beliefs that teachers already possess while they introduce new test systems into their current practice.

One obvious limitation of such studies is that since they simply focus on a narrow set of factors associated with testing itself, the authors and researchers are still not able to explain the nature of the washback phenomenon elaborately. Due to the narrow research focus, many assertions and statements made in these studies, though differing in wording, and overlap in meaning. In addition, although the issue of the different factors have been touched upon by a many researchers (Tan, 2008; Tavares & Hamp-Lyons, 2008) and begun to be explicitly and intensively dealt with in Turner (2008, 2009), additional data need to be collected to enable researchers to examine and address the issue more closely and extensively, and above all, to illustrate whether the findings from Canada, Hong Kong and South Africa apply to other contexts as well.

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

149

The findings of research on examination impact in the field of education, and on washback in the field of applied linguistics, however, have been mixed. Researchers in both fields have come to a similar conclusion that washback is a very complex phenomenon and that it is likely to be mediated by numerous factors such as contextual factors and stakeholders' beliefs. Despite the link between washback effects and mediating factors discussed in the literature, it is still not known exactly how washback works positively and negatively. Thus, the present research aimed at exploring in details the influence of washback on teachers and students in the context of Bangladesh educational reform through the standards-based assessment known as the HSC public examination. Previous worldwide studies on washback effects have revealed mixed results, indicating the complexity of washback. This interdisciplinary research attempted to explore the role of contextual factors and beliefs held by teachers and students in the process of washback, going beyond just identifying the nature of washback of the HSC public examination.

This chapter has started with an extensive overview of the washback research conducted both in the ESL and EFL context. First, it has examined the research on washback in language education and general education to clarify and summarise some basic concepts and perspectives related to the washback phenomenon. It has then offered a discussion about the washback studies carried out in the world context. After the brief introduction provided in Chapter One to the general context of the present study, and a broad set of theoretical and conceptual framework of washback outlined from multiple sources in Chapter Two, the next chapter presents the research methodology that was used to conduct the present research.

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

150

Chapter Four
Research Methodology

Research methodology refers to the systematic procedures and techniques used to carry out a study. This chapter describes the methodological procedures employed to collect and analyse data so as to answer the research questions posed in Chapter One. The chapter starts by presenting the overview of the research methodology. Then, it turns to the rationale for the methodology that has been applied in the study. After that, it describes the methods for data collection, the research design adopted, the instruments used, the participants involved, and the sampling. Finally, the data collection procedures and the process of data analysis are explained.

4.1 Research Methodology: An Overview

There is a general agreement that washback is a complex phenomenon. Many researchers call for empirical studies to explore the concept further. Alderson and Wall (1993) assert that the best way to identify washback is through a combination of teacher and/or student surveys and direct classroom observation. The literature on washback studies is increasing, and the methods employed for data gathering in these studies are diverse. Though the earlier studies on washback simply used a single data source, the later studies embraced multiple data sources. The methods employed in recent research studies tended to involve questionnaires, interviews, and classroom observations. (e.g,, Herman & Golan, 1991; Shohamy, 1992; Andrews and Fullilove, 1994; Qi, 2004, 2005; Cheng, 2001; Cheng, 2004; Davison, 2008; Shohamy, 1993; Tavares & Hamp-Lyons, 2008; Turner 2008, 2009; Urmston & Fang, 2008; Wall, 1999; Watanabe, 1996, 2004).). The review of washback studies also shows that there seemed to be no instruments that had been developed specifically for washback studies.

It is encouraging to note that during the last decade more and more researchers have expanded to look at issues of the context in order to capture the complexity of the washback phenomenon much more seriously, both theoretically

and empirically. It is also worth mentioning that adopting the mixed-method (MM) approach is the growing trend in current washback research.

4.1.1 Development of Washback Studies

Research is an ongoing process, and its design evolves over time. The research methods used vary from study to study. It should be noted that the methodologies utilized in washback studies have undergone a developmental change during the last couple of years. There has been an evolution in this field of research from the use of a single method or monomethod (e.g., survey methods) to the use of multiple methods or mixed methods (e.g., survey methods, in-depth interview, complemented by observations).

Between 1980 and 1990, little empirical research had been carried out to investigate the washback effect of examinations either in the field of general education or in the field of language education. Research design during that period was largely dominated by survey methods (usually interviews or written questionnaires), with observation being overlooked. Nevertheless, although the questionnaire data provided a great deal of information on the relationship between teaching, learning and testing, these data alone could hardly provide a clear and accurate portrayal of what was actually happening in the classroom.

It is widely acknowledged that the most substantive contribution in this area, which led to the popularization of the use of multiple methods, is the Sri Lankan Impact Study reported by Alderson and Wall (1993). Most important of all, it has motivated a substantial amount of evidence-based, observational washback research (Alderson & Hamp-Lyons, 1996; Burrow, 2004; Cheng, 1997, 1998; Read & Hayes, 2004; Qi, 2004; Shohamy et al., 1996; Turner, 2002, 2008, 2009; Watanabe, 1996a, 1996b, 2004b). The study of Alderson and Wall (1993) is a benchmark and the torchbearer in the field of the washback research. The research questions in washback research are the best answered with mixed-method research designs rather than with sole reliance on either the quantitative or the qualitative approach. Turner (2005, 2008, 2009) attested the importance of using multiple methods of data collection (a mixed-method design), and provided a good example of how rigorous washback research combining qualitative (QUAL) and quantitative (QUAN) methods could be designed.

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

152

4.1.2 Mixed Methods (MM) Research: Washback Study Context

As indicated above, a mixed-methods (MM) orientation has been embodied in the design characteristics of recent washback research (Bailey, 1999; Burrow, 2004; Cheng, 2001, 2003; Qi, 2004; Turner, 2005, 2008, 2009; Wall, 1999; Wall & Alderson, 1993; Watanabe, 2004). It was not until recently that the use of Mixed Methods Research (MMR) as a research design was articulated in researchers' explanations of their methodologies (Turner, 2008; 2009). Then, in what follows, the present researcher examines the theoretical groundings of the mixed methods research (MMR) design as well as some of the unique design features subsumed under it.

Teddlie and Tashakkori (2009) define a mixed methods (MM) study is one in which the researcher uses multiple methods of data collection and analysis. They argue that the mixed-method approach is underpinned by philosophies of pragmatism. Plenty of evidence shows that the MM approach has gained broad appeal in research from different disciplines (Creswell & Plano Clark, 2007; Greene et al., 1989; Johnson & Onwuegbuzie, 2004; Tashakkori & Teddlie, 1998; Teddlie & Tashakkori, 2009; Turner, 2005, 2008, 2009). Creswell and Plano Clark (2007) advocate conducting research along these lines, saying "the use of quantitative and qualitative approaches in combination provides a better understanding of research problems than either approach alone" (p.18). In light of these practical reasons provided by different researchers, it seems that there is a need to examine the theoretical grounding of this approach.

Greene (2007, p. 20) has noted, "The primary purpose of a study conducted with a mixed methods way of thinking is to better understand the complexity of the social phenomena being studied." As Johnson and Onwuegbuzie (2004, p.15) have stated, mixed methods research is "inclusive, pluralistic and complementary…[it] take[s] an eclectic approach to method selection and the thinking about and conduct of research." For them, what is primordial is the research question; research methods are solutions that work to answer the research question(s) best. It is interesting to note that their argument is reinforced by Creswell and Plano Clark (2007); they assert that investigators may view MMR strictly as a ''method,'' thus allowing

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

153

researchers to choose any method from different schools of methodology based on diverse philosophical assumptions.

The importance of the MM approach lies in that it allows researchers to mix aspects of the qualitative and quantitative paradigms at all or many methodological steps in the design (Creswell, 1994; Creswell & Plano Clark, 2007). Patton (1990) has conceptualized "methodological mixes" saying that different methods: QUAN (quantitative), and QUAL (qualitative) could be combined across three stages: design, measurement (QUAL data or QUAN data), and analysis (content or statistical). Allwright and Bailey (1991) extend Patton's (1990) conceptualization saying that various combinations of quantitative and qualitative data collection and analysis are possible. Creswell and Plano Clark (2007) provide a more elaborate definition of four major types of MM design:

a. *Triangulation design-* **A** triangulation design refers to the collection of qualitative and quantitative data simultaneously to understand a problem;

b. *Embedded design-* An embedded design means using qualitative data in an experiment or correlational study;

c. *Explanatory design-* An explanatory design explains quantitative results with qualitative data; and

d. *Exploratory design- An* exploratory design uses qualitative data and analysis in an exploratory function towards developing a quantitative instrument.

It is worthwhile to note that the four types of designs address different objectives. They can serve as a foundation for conceptualizing how to design and conduct feasible MMR. Research in washback studies (e.g. Wall & Alderson, 1993; Cheng, 1997, 1998; Turner, 2002, 2005, 2008, 2009; Wall, 1999) demonstrates that all of the MM designs used triangulation techniques. Such designs stress the importance and predominance of the research question over considerations of either method or paradigm (e.g., the worldview that is supposed to underlie that method). Subsumed under the MM approach is an array of methods combining both quantitative and qualitative research: observations, interviews, document reviews, questionnaires and so on.

Creswell (2009), Creswell and Plano Clark (2007), Greene (2007), Johnson and Onwuegbuzie (2004), and Teddlie and Tashakkori (2009) suggest that MMR

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

154

produces better outcomes than mono-method research. According to these researchers, MMR has the potential to reduce some of the problems associated with single methods. From their perspective, by utilizing quantitative and qualitative techniques within the same framework, MMR can incorporate the strengths of both methodologies. In light of the above perspective, in order to examine and understand the phenomenon in questions, it is necessary to draw upon both types of data (QUAL and QUAN).

The MMR places on tailoring methods to research questions. As was put by Johnson and Onwuegbuzie (2004), research approaches should be mixed in ways that offer the best opportunities for answering important research questions. Based on their explanation, MMR does not dictate the choice of data collection methods. Rather it allows the procedures for conducting research to be dictated by the research question and the context of the study. One of the salient strengths of the MM approach lies is that it allows researchers to mix aspects of the qualitative and quantitative paradigms at all or many methodological steps in the design (Creswell, 1994; Creswell & Plano Clark, 2007).

For its strengths, the MM approach is becoming more and more popular with researchers in the domain of washback research. Except for Watanabe (1996) and Alderson and Hamp-Lyons (1996), the majority of washback studies have embraced this approach. In fact, the Sri Lanka Impact Study as well as the study by Turner (2002, 2006, 2008, 2009) has demonstrated a successful combination of survey research and QUAL procedures. Turner (2005, 2008) clearly states that the research design and analytic procedures of her study have been informed by the principles of the MM approach. Cheng's (1997, 1998) longitudinal study relies heavily on QUAL methods such as observations, interviews, and document analysis, but incorporating a complementary QUAN component (e.g., questionnaires).

From a methodological perspective, Cheng (1997, 1999, 2000) forcefully argues that the complex washback phenomenon necessitates the use of both QUAL and QUAN research methodology. Her argument is strongly supported by Watanabe (1996) and Chen (2002) who also strongly believe that QUAL and QUAN methods can be profitably used together in the study of washback. Because of the role and importance of the mixed-method approach in the washback studies, the present researcher adopted the appropriate methods which most washback research

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

155

prescribed. Tsagari (2007) has listed 29 empirical researches and their methods which include questionnaire, observation, interviews, and analysis of documents. Some researchers also use test scores, test analysis, and case studies. The reason why this method has largely been favoured by washback researchers (Alderson & Hamp-Lyon, 1996; Cheng, 2003; Watanabe, 1996a, 1996b, 2004b; Turner, 2002) is that it is held to be able to produce a set of information-rich data (Cheng, 2003; Watanabe, 2004a).

According to Greene (2007), MMR, with its emphasis on holistic, richly detailed descriptions and analyses of teaching behaviours and the multilevel contexts in which those behaviours are nurtured, is best suited for capturing the complexity of the social phenomenon being studied (Greene, 2007). Meanwhile, as noted by Turner (2006, 2007), the MMR has the potential to "help respond to certain types of questions, especially those having to do with classroom contexts" (2009, p.108). In this regard, this approach seems to be best suited for the present research purpose.

4.2 Research Methodology for the Present Study

In the overview of research methodology, an introduction is provided to the MM design as well as the rationale for utilising the MM approach in washback research. Owing to the focus of the study, the decision was made to conduct this study utilising such an approach. The present researcher combined aspects of quantitative and qualitative methods in the stages of data collection and data analysis. Qualitative data collection mainly involved in-depth interviews, classroom observations, and analysis of HSC examination related papers. The quantitative data collection consisted of the completion of a questionnaire. With respect to the choice of research methods, an essential first step to be taken involved an examination of all relevant and available documents related to the HSC examination. In this study, the present researcher conducted an intensive review and analysis of the documents pertaining to the HSC syllabus and curriculum and its objectives targeted by the National Curriculum and Textbook Board (NCTB) reflecting the EFL education intentions, the HSC question papers in English, and the textbooks used at this stage. The quantitative data collection consisted of the completion of a questionnaire.

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

156

The MM approach was deemed an appropriate avenue because of its strength for addressing the research questions of the present study. This approach was chosen based on three aspects of the study: the type of problem to be addressed, the goal of the study, and the nature of the data. The purpose of adopting this approach was to devise a solid research design that might maximize the possibility of addressing the research questions thoroughly. The researcher conducted the study in a scientific manner, and proceeded step-by-step applying the following methodology (Table-4.1):

Table 4.1: Research design of the present study

Research Design Methodology Considerations **[Mixed Methods Research (MMR) Approach]**	
Quantitative Method: Questionnaire Design 1. **Teacher survey:** 2. **Student survey:** • Five-Grade Likert Scale (Likert,1932) used	**Qualitative Method:** 1. **Classroom Observation** 2. **In-depth Interview** (EFL teachers, EFL examiners, and curriculum specialists) 3. **Analysis of the HSC examination related documents** (e.g. HSC syllabus and curriculum, textbook, question papers, and the answer scripts).

↓

	Data Collection →		**Data Analysis →**
Data Collection	• Pilot Study • Baseline Data		• Phase- I
	• Teacher and Student Questionnaire Surveys		• Phase- II
	• Classroom Observations and Interviews		• Phase- III
	• In-depth Interviews		• Phase- IV
Data Analysis	Software/Tools Used	• Computer Package SPSS 18.0 for Widows • Microsoft Excels	
	Qualitative analysis	• Comparative method, Inductive logic/analysis Categorize emerging themes, • Developing theory • Organizing the data • Reporting the outcomes	
	Quantitative analysis	• Frequency counts • Descriptive statistics (e.g., frequency counts, means, standard deviations, etc.) • Inferential Statistics	

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

157

The above mixed methods (MM) research design was used for collecting both quantitative and qualitative data; it means the data collection process tried to ensure that there could be comparisons made between and within all designated levels and categories across data collection periods and also across the different kinds of data whenever possible. It was believed that the combination of these research methods would allow the present researcher to examine the washback on the EFL teaching and learning from many different angles.

The principle of triangulation is particularly appropriate when investigating complex issues such as washback. For the present study, the data triangulation was achieved by having different sets of data cross-checked. In addition, other standards such as "persistent observation", "thick description of the content", and "explicit emphasis on research question(s)" were also taken into account in the present study.

4.2.1 Triangulation of the Present Study

The principle of triangulation is particularly appropriate when investigating complex issues such as washback. Triangulation is "the use of two or more methods of data collection in the study of some aspect of human behaviour" (Cohen and Manion 1994, p. 233). To elaborate on this point, Brown (2000), quoting Rossman and Wilson (1985), presents the view that "Data from different sources can be used to corroborate, elaborate, or illuminate the research question" (2001, p. 227). Denzin (1978) uses the term triangulation to define the combination of the data collection sources. Regarding triangulation, Glesene and Peskin (1992) states that the data collected from the multiple sources enhance the trustworthiness and credibility, thereby increasing confidence in research findings. Marshal and Rossman (1989) argue that using a combination of data sources increase the validity of the findings. There are essentially four types of triangulation:

The first one is *data triangulation,* in which data from more than one source are brought to bear in answering a research question (e.g., the data from teachers, language learners, and inspectors in the study by Shohamy et al., 1996).

Second, *investigator* (or *researcher) triangulation* refers to using more than one person to collect and/or analyse the data.

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

158

Third, in *theory triangulation* more than one theory is used to generate the research questions and/or interpret the findings.

Finally, in *methodological* (or *technique) triangulation* more than one procedure is used for eliciting data, for instance, Wall and Alderson's (1993) use of interviews and classroom observations.

Brown (2002) observes that triangulation must be carefully planned; otherwise there is no guarantee of the validity of the results. He reminds researchers of the importance of acknowledging any preconceptions or biases that might affect their choice of data. Here, for the present study, two forms of triangulation were employed – (i) data triangulation, where data were collected from a number of sources ; (ii) and methodological triangulation, where different techniques were used to elicit the data. This study may be considered a pilot one for future washback studies as it was designed to investigate and learn techniques to explore washback in classrooms, refine classroom washback observation instruments, identify potential differences and variables which might indicate or effect washback, identify useful statistical tools, evaluate the time frames and sample sizes for such investigations, and indicate the necessary scope of future washback investigations.

Wall and Alderson's (1993) study in Sri Lanka provides an excellent example of investigator triangulation and methodological triangulation. Investigator triangulation is illustrated by the fact that "seven Sri Lankan teachers based in five different parts of the country agreed to act as observers" (p. 49), and went through a three-month training programme to prepare for this role. The resulting data included "questionnaires, interviews, materials analysis, and most importantly, observations of classroom teaching" (ibid., p. 44). Therefore, triangulation should be incorporated as a methodological cornerstone in any serious investigation of washback.

From the *data triangulation* point of view, the present researcher explored multiple data sources: language teachers, language learners, examiners, curriculum designers and policy makers. In *methodological* (or *technique) triangulation* consideration, several instruments such as questionnaire survey, interviews, classroom observation, and analysis of exam related documents were used in the present study to obtain required data. Multiple data sources and using a number of instruments helped the researcher have more authentic and reliable data.

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

159

4.2.2 Sampling of the Study

The population is a set of people or entities to which findings are to be generalised. The population must be defined explicitly before a sample is taken. A sample is a subject chosen from a population for investigation. However, random samples are always strongly preferred as only random samples permit statistical inference. In random sampling, all populations have the same chance to be selected, and can be calculated in a study. A random sample is one chosen by a method involving an unpredictable component. Random sampling can also refer to taking a number of independent observations from the same probability distribution, without involving any real population. That is, there is no way to assess the validity of results of non-random samples. The present study used "Simple Random Sampling" while selecting the respondents.

Morris (1996) suggests that the advantage of random sampling is that it is easy to apply when a big population is involved (p.17). Robert (1997) opines that random sampling is inexpensive and less troublesome (p.103). Agresti (1983) suggests that a sample must be large to give a good representation (p.23). Cochran's formula (Cochran, 1977, Wang, 2010) was used to determine an appropriate sample size of students and teachers. The target populations/subjects for the present study were higher secondary students, English language teachers, examiners, policy makers, and the curriculum specialists.

4.2.2.1 Subjects

In the last two decades, most researchers used mostly two types of respondents in washback studies: teachers, and students (e.g. Alderson and Wall 1993, 1996, Turner, 2001, 2002, 2005, 2008, 2009; Shohamy, 1993,1996); the other types of respondents frequently used were: policy makers, curriculum designers, administrators, testers, test developers, textbook writers (e.g. Cheng, 1997, 1998, 1999, 2001, 2003, 2004). Like the previous washback studies in different countries and contexts, the samples of the present study consisted of teachers, students, and some other stakeholders (examiners, policy makers and the curriculum specialists).

Based upon the consideration of statistical power, three different formulas, developed respectively by Cochran (1977), Krejcie and Morgan (1970), and

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

160

Scheaffer et al. (1996), were compared to decide on an appropriate sample size. The respondents were selected both from urban (50%) and rural (50%) colleges. A questionnaire survey was conducted among 500 higher secondary 2nd year students and 125 teachers teaching English to the same students. The other participants were 4 EFL examiners and 3 curriculum specialists; the classroom observation participants were 10 English teachers and their students. Among the factors that can mediate the washback effect is the teacher (Wall, 1996) and her/his perceptions about the examination, its nature, purposes, relevance in the context, etc. The participants had the following characteristics:

Firstly, the teachers were currently teaching English at higher secondary level in Bangladesh. So, they could provide the needed information related to the research topic. Some of them were EFL examiners of English subject of the HSC public examination.

Secondly, the students had been studying English at the HSC level; they completed 12 years of schooling with English as a compulsory subject.

Thirdly, the interviewed curriculum specialists had been working in the NCTB. They were selected to elicit information on various issues of EFL testing, teaching, and curriculum objectives; and they were interviewed through semi-structured questionnaires.

Finally, the participants were volunteers, and willing to respond the topic without force. The names of teachers, students, examiners and curriculum specialists are anonymous.

4.2.2.1.1 Research Sites and Selection of Participants

The participants for the study can be divided into 3 broad categories. The first category of participants took part in the questionnaire surveys. The second category of participants took part in the classroom observations. The third category of participants took part in the semi-structured interviews. The participants were chosen on the basis of their potential for yielding data which could reveal participants' perceptions in general. The research sites included 18 higher secondary government and non-government colleges under 8 districts and the

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

161

National Curriculum and Textbook Board (NCTB). The details about these categories of participants and locations are displayed in Table 4.2.

Table 4.2: Research sites and participants

Name of Districts Participated in the Study		No. of Research Sites	Questionnaire Survey Participants	Classroom Observation Participants	Interview Participants		
					EFL Teachers	EFL Examiners	Curriculum Specialists
1	Dhaka	04	√	√	√	√	X
2	Gazipur	02	√	√	√	√	X
3	Narayangonj	02	√	√	X	X	X
4	Tangail	04	√	√	√	√	X
5	Narshindi	02	√	X	X	X	X
6	Jamalpur	02	√	X	X	X	X
7	Manikgonj	01	√	X	X	X	X
8	Mymensingh	01	√	√	X	X	X
9	NCTB	01	X	X	X	X	√

The present researcher ensured that the participants came from colleges of various geographical locations: large/small cities, rural/urban, north/south. In addition, the selection of participants was also largely based on practical considerations and participants' willingness and interest to discuss specific issues on the HSC examination and its influence on teaching and learning English as a foreign language.

4.2.2.1.2 Questionnaire Participants

The questionnaire survey participants consisted of 500 higher secondary students and 125 EFL teachers selected from 18 higher secondary colleges. They were believed to represent the largest population of the higher secondary education in Bangladesh. The survey was administered during April 2010 - July 2010. The researcher himself administered the survey in the selected higher secondary colleges.

4.2.2.1.3 Classroom Observation Participants

The researcher observed 10 EFL classes in 10 different colleges in Bangladesh. Upon selection of teachers for the study, specific class sessions were chosen for observation. The classes were selected on the basis of the lesson scheduled for that day and its relationship to the HSC examinations. During the observation, 10 EFL teachers and their 511 students were observed; and all of the 10

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

162

observed teachers and 355 observed students participated in the questionnaire surveys.

4.2.2.1.4 In-depth Interview Participants

Taylor and Bogdan (1998) specify that interviews are useful in research contexts where the researcher's interests are fairly well-defined. The interview participants, as already mentioned, were 6 EFL teachers, 4 EFL examiners, and 3 curriculum specialists. They were interviewed through semi-structured questionnaires.

4.2.2.2. Instrumentation

A good number of previous washback studies elicited data employing questionnaires, interviews, testing measures and classroom observations (e. g. Wall and Alderson, 1996; Herman & Golan, 1991; Shohamy, 1992; Andrews and Fullilove, 1994; Qi, 2004; Tsagari, 2007, 2009; Wang, 2010). In keeping with the general approach outlined above, several instruments were used in the present study. As indicated in the previous sections, the present researcher applied mixed methods research (MMR) approach combining aspects of quantitative and qualitative methods in the stages of data collection and data analysis.

The present researcher used in-depth interviews, classroom observations and analysis of HSC examination related material to elicit qualitative data, and conducted a questionnaire survey for students and teachers to obtain quantitative data, which provided ample insights into the relationship between teachers' perceptions of teaching contents and public examinations. The interviews were semi-structured, conducted in a systematic and consistent order, which nevertheless allowed the present researcher sufficient freedom to probe far beyond the answers to the prepared questions (Bogdan & Biklen, 1998).

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

163

4.2.2.2.1 Questionnaire Survey

It is proven that different methods may perform different functions in different studies. In general, many researchers have used surveys (e.g., questionnaires) not only to gather information about participants' characteristics but also to uncover the opinions and attitudes of the participants about washback as well as their views and perspectives on language teaching and learning (Cheng, 2004; Qi, 2005; Turner, 2005, 2008, 2009; Watanabe, 1996a). Cheng (2004) views questionnaires as being able to provide a general picture of how teachers and students react. The strength of survey, based on Watanabe (1996a), lies in that it can detect and explain the reasons behind teachers' behaviours in classrooms. Similarly, but more explicitly, Qi (2005) states that the goal of employing questionnaires in her study is to find out how far the interview results can be applied to a larger group of participants.

If properly designed and implemented, surveys can be an efficient and accurate means of determining information about a given population. The results from questionnaires can be provided relatively quickly; and depending on the sample size and methodology chosen, they are relatively inexpensive. For this reason, the questionnaire has become one of the most popular methods of data collection in education research. It is generally considered an efficient (cheap and fast) method of gathering information from a large number of respondents. Another advantage of using a questionnaire is its high reliability. The questionnaire survey technique involves the collection of primary data about the subjects, usually by selecting a representative sample of the population or universe under a study through the use of a questionnaire. It is very popular since many different types of information can be collected: attitudinal, motivational, behavioural and perceptive aspects. It allows for standardisation and uniformity both in the questions asked and in the method of approaching subjects, making it far easier to compare and contrast answers by the respondent group.

The present researcher used two types of questionnaires: student questionnaire, and teacher questionnaire, to collect quantitative data. Two separate sets of questionnaires were constructed for both students and teachers. The questionnaires were in Five-point Likert scales (Likert, 1932) ranging from "strongly agree" to "strongly disagree" (strongly agree; agree; no opinion; disagree;

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.　　<Pages 288-786>

164

strongly disagree). The 45-item questionnaires were constructed on a number of domains which were affected or influenced by washback such as the syllabus and curriculum, material, teaching method, feelings and attitudes, teaching-learning strategies, and learning outcomes. The researcher developed the questionnaires following the model of Mizutani (2009), Hayes (2003), Chen (2002), Al-Jamal and Ghadi (2008), Tsagari (2007), Green (2007), and Wang (2010). The above models of questionnaires were followed because they dealt with major areas of language testing and teaching, and they were relevant to the present study; their models proved to be appropriate for investigating washback of high-stakes examination, as the HSC examination in Bangladesh. The models of Turner (2002, 2005, 2008, 2009), Latimer (2009), Jin (2010), and Hsu (2009) were consulted for proving the validity, reliability and practicality of the questionnaires.

The researcher prepared the questionnaires in the light of the research questions and the research objectives of the present study. The questions explored the particular washback of the HSC examination and the EFL teaching and learning topics. The items of the questionnaires were straightforward, and the linguistic nature of each question was relatively easy and simple.

The questionnaires were distributed to the respondents directly by the researcher; and the participation was voluntary, and the questionnaires were anonymous. A pilot test was conducted to check the reliability, validity and appropriateness of the questions. Item suitability, item relevance, clarity, and language diction were verified through the pilot test. In line with the recommendations of Wang (2010), the present researcher worked through each of the following areas in sequence: determination of primary and subsidiary aims of the survey, determination of the target population, determination of the approach to recording and analyzing response data, consideration of ethical protocols, production of draft, trailing of the draft, revision of the draft, conducting the survey, and analysing the results.

In the present study, the closed format questions were chosen. They have the following advantages:

    a. Closed format questions have many advantages in respect of time and money.

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

165

b. By restricting the answer set, it is easy to calculate percentages and other statistical data over the whole group, or over any sub-group of participants.

c. The SPSS makes it possible to administer, tabulate and perform analysis in a relatively shorter period of time.

d. Closed format questions allow the researcher to filter out useless or extreme answers that might occur in an open format question.

e. The quality of a questionnaire can be judged by three major standards: (1) validity, (2) reliability, and (3) practicality.

The previous studies show that rather than being a direct and automatic effect, washback is a complex phenomenon. Furthermore, washback exists in a variety of teaching and learning areas (e.g. curriculums, methods of teaching, classroom assessment, student learning, feelings and attitudes of teachers and students). Therefore, the present 45-item questionnaire dealt with the questions related to the areas of the syllabus and curriculum, teaching method, teaching strategies, teachers' and students' perceptions of and beliefs in the examination, textbooks and materials, task and activities, etc.

4.2.2.2.1.1 Student Questionnaire

The student questionnaire consisted of 45 items covering 6 areas relating to examining the washback of the HSC examination on teaching and learning English as a foreign language such as general comments and their perceptions on the objectives of the syllabus and curriculum, effects on teaching materials, effects on teaching methods, what the learners wanted to learn, what perception and attitudes of the students had as to the public examination, how they practiced EFL skills and linguistics elements, etc.

The student questionnaire (Appendix-1A) was structured in six sections. Section One (from Q1- Q7) aimed to solicit questions about the syllabus and curriculum such as curriculum objectives, teaching the items in the syllabus, skipping items and lessons, etc. Section Two (from Q8-Q17) consisted of a set of questions related to the textbook *English for Today* (EFT) and other materials used in the class. Section Three included questions (from Q18 – Q26) concerned with teaching methods and classroom behaviours. The questions in Section Four (from

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

166

Q27- Q32) were about the classroom tasks and activities which usually took place in the class. Section Five included questions (from Q33- Q37) related to practicing the different skills and linguistic elements of EFL. The last section (Q38-Q45) consisted of questions as to the students' attitudes, beliefs, and perception towards the HSC examination.

The questions were closed-ended items in different issues. It included Likert-type questions. The scale used in the Likert-type questions ranged from 'strongly agree' to 'strongly disagree' (strongly agree=5, agree=4, no opinion=3 disagree =2, and strongly disagree=1). The survey was conducted from April to July 2010. The completion of the questionnaire took approximately 30 minutes. The questionnaire covered the following domains (Table 4.3) of the EFL teaching and testing.

Table 4.3: Taxonomy of student questionnaire

SL	Components	No. of Items	Question No.
1	Syllabus and Curriculum	7 items	Q1- Q7
2	Textbooks and Materials	10 items	Q8- Q17
3	Teaching Methods	9 items	Q18-Q26
4	Tasks and Classroom Activities	5 items	Q27-Q32
5	Language Skills and Elements	5 items	Q33-Q37
6	Students' Attitudes and Perception related the test and teaching	7 items	Q38-Q-45

The themes of the student questionnaire were based on the issues that were used in many studies to examine the complexity and dimension of the washback on the EFL/ESL teaching and learning in different contexts (e.g., Hayes, 2003; Al-Jamal. and Ghadi, 2008; Wang, 2010; Tan, 2008; Hsu, 2009, Alderson and Wall, 1993,1996; Saif, 1999; Satomi, 2009). Therefore, the reliability, validity, authenticity, and practicality of the present questionnaire were sufficiently maintained from the start. Besides, the present researcher conducted a pilot study twice upon the same students. It was the test-retest method to compute the reliability, validly, and practicality of the instrument. The questionnaire was first administered on 20 higher secondary students (not included in the sample of the study), and then administered once again on the same group three weeks later. Spearman's coefficient of correlation formula (1947) was used in order to find out the reliability coefficient; and the ratings were considered to be sufficient for the purpose of applying the questionnaire, which was 0.93 for the first time of the study, and 0.91 for the second time (a perfect positive correlation):

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

167

$$r_s = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

The student questionnaire was highly valid with regard to the content, construct, and criterion; the questionnaire dealt with the questions that directly matched the investigation of the study. It is crucially important that a questionnaire must be practical to be administered. Practicality involves the cost and convenience of the test. The student questionnaire of the present study had high level practicality because it was relatively cheap to produce (economic); it took nearly 30 minutes to answer all the questions; and the analysis of the results could be described by descriptive statistics.

4.2.2.2.1.2 Teacher Questionnaire

It is strongly assumed that questionnaires are versatile, allowing the collection of data through the use of open or closed format questions. The teacher instrument was 45-item questionnaire (Appendix-1B) prepared with the same mechanism as followed in the student questionnaire. The questionnaire covered the issues that were used by many previous studies to examine the complexity of the washback on EFL/ESL teaching and learning in different contexts (e.g., Hayes, 2003; Al-Jamal and Ghadi, 2008; Wang, 2010; Tan, 2008; Hsu, 2009, Alderson and Wall, 1993, 1996; Saif, 1999; Mizutani, 2009). Thus, certain degrees of validity such as construct, predictive, and content can be assumed from the formation level of the questionnaire.

The present researcher tested the reliability and validity of the questionnaire in a number of ways: conducting a pilot study in the form of test - reset, checking by the supervisor of the researcher, reviewing by the senior researchers. Therefore, the reliability, validity, authenticity and practicality of the questionnaire were made confirmed. Besides, the pilot study was conducted to compute the reliability of the instrument. The questionnaire was first administrated on 10 higher secondary English language teachers (not included in the sample of the study), and then administrated again on the same group two weeks later. Like the student questionnaire, Spearman's coefficient of correlation formula was used in order to find out the reliability coefficient of the teacher questionnaire, which was

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

168

0.91 for the first time of the study, and 0.89 for the second time (a perfect positive correlation).

The teacher questionnaire followed the model of student questionnaire which was structured in six sections. The first section aimed at soliciting questions about the syllabus and curriculum such as curriculum objectives, teaching the syllabus, skipping items and lessons, etc. The second section consisted of a set of questions related to the textbook *English for Today* and other materials used in the class. The third section included questions on teaching methods and classroom behaviours. The questions in the fourth section were on the classroom tasks and activities that usually took place in the class. The fifth section included the questions on skills and linguistic elements of EFL usually practiced by them. The last section consisted of questions on the attitudes, beliefs, and perception towards the HSC examination. The teacher questionnaire dealt with the following areas of EFL teaching and testing (Table 4.4).

Table 4.4: Taxonomy of teacher questionnaire

SL	Components	No. of items	Question No.
1	Syllabus and Curriculum	7 items	Q1- Q7
2	Textbooks and Materials	10 items	Q8- Q17
3	Teaching Methods	9 items	Q18-Q26
4	Tasks and Classroom Activities	5 items	Q27-Q32
5	Language Skills and Elements	5 items	Q33-Q37
6	Students' Attitudes and Perception related the test and teaching	7 items	Q38-Q-45

The questions were closed-ended on different issues. The scale used in the Likert-type questions ranges from 'strongly agree' to 'strongly disagree' (strongly agree=5, agree=4, no opinion=3 disagree =2, and strongly disagree=1). The final survey was conducted during April - July 2010. The completion of the questionnaire took approximately 30 minutes.

4.2.2.2.2 Classroom Observation

Observation is a primary method of collecting data by human, mechanical, electrical or electronic means. The observation sessions are carried out to address research questions to recapitulate, speculated on the extent to which teachers are

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

169

influenced by test contents. According to Wall and Alderson (1993), the perceived value of classroom observation is that it allows researchers to have more direct access to the teachers' behaviours and interaction patterns in the classroom. In their words, it can help determine what teachers teach, and how. Moreover, it eliminates the need to ask individuals about their behaviours or tendencies which are sometimes not reliable (e.g., Alderson & Wall, 1993; Cheng, 1997; Wall & Alderson, 1993; Shohamy, 1993; Turner, 2005).

For the present study, the amount of communicative methodology that teachers actually implemented at the classroom level was observed; on average 51 students were found present in each EFL classes during observation. The observation schedules "Communicative Orientation to Language Teaching (COLT)" scheme (Appendix-2A) and "Modified University of Cambridge Observation Scheme (UCOS)" (Appendix-2B) were used. In addition, a self-made checklist for the classroom observation was prepared to elicit additional information that was not in the two schedules above. The observation checklist included examination related classroom activities and the teachers' personality issues (Appendix-2C). Observation techniques can be part of qualitative research as well as quantitative research techniques.

The main purpose of the observation was to find out whether the HSC examination in English could foster an impact on EFL classroom teaching and learning. Meanwhile, it was hoped that conducting classroom observations might help determine whether teachers' accounts of their beliefs, their understanding of ELT methodologies as well as their attitudes towards washback conformed to their classroom behaviours.

One distinct advantage of the observation technique is that it records actual behaviours of the teachers. Indeed, sometimes their actual recorded behaviour can be compared to their statements, to check the validity of their responses. Especially, when dealing with behaviour that might be subject to certain social pressure (for example, people deem themselves to be tolerant when their actual behaviour may be much less so) or conditioned responses (for example, teachers say they value communicative competence, but will apply the grammar-translation method and isolated vocabulary teaching), the observation technique can provide greater insights

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

170

than an actual survey technique. The present researcher applied a semi-structured observation approach and followed the steps below (Figure 4.1):

Figure 4.1: The development model of the observation checklist
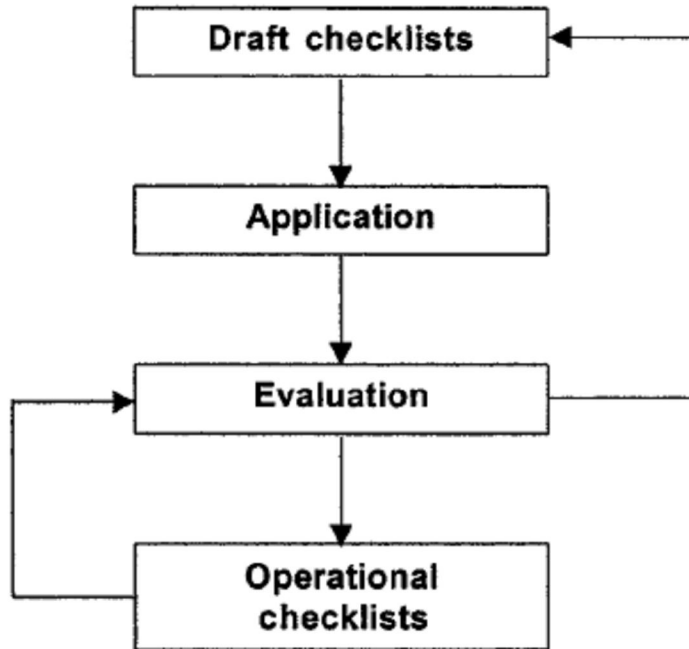
For this study, the present researcher observed the 10 higher secondary EFL classes taught by the English teachers (who also participated in the questionnaire survey), and recorded classroom observations activities while observing. Along with the COLT and UCOS, the present researcher conducted semi-structured observation covering a number of areas to answer the research questions. The present researcher applied this method because it had offered an effective way to accurately record the maximum amount of information describing what occurred in the classroom, and had been used successfully by the researcher in a multitude of classroom observations. The format of the observations sheet allowed the researcher to record everything said by the teacher and the students, with dedicated columns for each. The researcher also recorded the time for each event in the classroom, which enabled him to calculate the percentage of class time spent on each activity, and then calculate how much time teachers spent on specific topics overall.

Reliability of the observations was checked against observations recorded by independent researchers (e.g., Wang, 2010; Fournier-Kowaleski, 2005; Hayes, 2003). Classroom observations were carried out on a small scale among those

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

171

teachers who were willing to be observed. As the observation procedure was still in progress, the only changes observed lay in the different language activities teachers employed in their teaching. Ten teachers agreed to participate in the observation. They were three female and seven male; all were qualified teachers. This group of teachers was not meant to be representative of all the teachers of English in Bangladeshi higher secondary colleges. The teachers were selected using purposive sampling (Patton, 1990), and the main purpose was to select teachers based on whether they could provide a rich variety of information about classroom teaching and learning activities in the classrooms in relation to the HSC examination in EFL.

4.2.2.2.2.1 Rationale for the Classroom Observation Study

Classroom observation views the classroom as a place where interactions of various kinds take place, affording learners opportunities to acquire. To reiterate, this study dealt with possible impacts that the implementation of the EFL test requirement might bring about in classroom teaching and learning in Bangladesh over a period of time. Therefore, observation was an essential instrument. There are essentially two different approaches to classroom observation: structured observation and unstructured observation. Highly structured observation involves going into the classroom with a specific purpose and with an observation schedule with pre-determined categories, and is usually linked with the production of quantitative data and the use of statistical analyses (Denscombe, 2007).

With the observation schedule, the observer records what participants do, as distinct from what they say they do. Because the observer is not required to make inferences during the data collection process, the schedule effectively eliminates any bias from the observer, and appears to produce objective data. Therefore, with structured observation, it is possible to achieve high levels of inter-observer reliability, in the sense that two or more observers using the same schedule should record very similar data. Unstructured observation, on the other hand, is less clear on what it is looking for, and usually requires the researcher to observe first what is taking place before deciding on its significance for the research study. Thus it involves recording detailed field notes, and produces qualitative data. It allows observers to gain rich insights into the situation, and is suited to dealing with complex realities.

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

172

The weaknesses of the two approaches have been debated (Allwright & Bailey, 1991; Denscombe, 2007). Structured observation records what happens, but not why it happens. It does not deal with the intentions that motivated the behaviour. In addition, unless a researcher is very clear about what exactly to observe and designs a well-tested observation scheme, the subtleties of the situation can easily be ignored. The data from an unstructured observation usually relies heavily on the researcher's inferences and detailed field notes in a particular context, which create problems with respect to the reliability and representativeness of the data. As the two approaches to classroom observation have their individual advantages and disadvantages, they would better be used complementarily rather than exclusively.

The investigation into the washback effect of English proficiency tests on teaching and learning presented a complicated research situation. It was clear from the start that there would be many intervening factors that interacted in teaching and learning as a result of the implementation of English proficiency tests. This seemed to require a combined approach using both observation approaches, resulting in what might be called a semi-structured observation. Therefore, semi-structured observation was best suited for the present research.

4.2.2.2.2.2. Observation Schedule

Observation has long been accepted as an important feature in language education and supervision, but for the past two decades, it has become established as the key process in language classroom research as well. The present researcher conducted classroom observation as one of the major instruments for obtaining relevant data. For this, two observation schedules were designed and applied following the Communicative Orientation to Language Teaching (COLT) Scheme, and University of Cambridge Classroom Observation Schedule (UCOS). The schedules were applied based on the analysis of the data derived from the questionnaires and document analysis.

4.2.2.2.2.3 Use of the COLT, Part- A, and UCOS

The present researcher mainly used the COLT (Part-A). The UCOS was also used at times as a complement to the COLT when it was necessary. One of the

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.　　　<Pages 288-786>

173

advantages of COLT (Part-A) is that it can be adapted to different contexts. In this study, Part A of the COLT scheme (Appendix- 2A) was used in its original version to allow the researcher to become familiar with the instrument and to determine its usefulness in this context. The instrument COLT (Part-A) was designed to be completed with the observer coding the classroom events as they occur. In this study, detailed notes of the activities and episodes were taken during the lessons. Part B of COLT, which focuses on the communicative features of classrooms, was not used as this level of linguistic analysis was beyond the scope of the study.

The classroom observation schedule UCOS was used as the second option of the classroom analysis. The instrument contained lists of text-types used in the classroom and a range of task types according to skills. It also identified teacher initiated, exam-related activities as well as grammar and vocabulary activities. On occasions, when activities observed were not adequately represented in by the categories in the original form, the instrument (Appendix -2B) was modified so as to reflect what occurred in the class.

Several significant activities were also observed through a self-made checklist (Appendix-2C) during the lessons, which were not specifically identified by either COLT or the UCOS. These were recorded and analysed separately. For example, features such as the teacher giving the students information about the examination or discussing test-taking strategies was specific to the type of class being studied. Instances of the teacher working with individuals or small groups were not adequately reflected within the COLT analysis, which focused on the primary classroom activity. Additionally, the study required a more detailed analysis of classroom materials than COLT could provide in its original form. In intensive courses, such as the ones observed, class time was limited; therefore the amount and type of homework given to each group of students was also recorded. Finally, the instances of laughter in each of the lessons were recorded in order to gain some indication of the atmosphere in each lesson, as was done by Alderson and Hamp Lyons (1996) and Watanabe (1996b) in their washback studies.

The objectives of the syllabus and curriculum generated, and the literature review formed the basis of the observations. The instrument was designed to record the following aspects of information:

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

174

1) **Observation Outline**: The researcher checked on student-centered activities (e.g., pair-work, group work, individual work, role-play), and counted the percentage of class time spent on teacher-centered activities (e.g., teacher lecturing to the whole class without interactions with students– teacher presentations, explanations of sentences, reading aloud, translations, etc.). The purpose of exploring classroom organization patterns in teachers' instructional process was to find out who was holding the floor in the classroom.

2) **Teachers' Instruction Dimensions**: The researcher counted the frequency of explaining language points with a focus on language forms (e.g., explanation of sentence structures, rote practice and mechanical grammar exercises; explanation of vocabulary in a decontextualized manner). He also calculated the frequency of involving students in meaning-based activities (e.g., discussion, role-play, comprehension exercises at the discourse-level, etc.). This was designed to evaluate whether the lessons delivered by the teachers were form-focused or meaning-focused, and to what extent the teachers' instruction was communicatively oriented.

3) **Relevance to the Test**: The present researcher documented and analysed use of class time spent on aural/oral aspects of English (e.g., listening practice, oral practice at the discourse level encouraged by the NCTB); frequency of giving information or advice about the HSC examination in English or test-taking strategies. This section was devised to discern whether and to what extent the teachers' instruction was related to the HSC examination.

4) **Medium of Instruction**: The researcher observed whether the teachers used English/Bengali/half English/half Bengali/ in the class as a medium of instruction. This was designed to learn about the language used by the teachers in their instruction, and teaching method/ approach they applied.

5) **Teaching Materials:** The researcher observed and recorded the types of materials used in the class: textbooks, test-related materials (e.g., the past examination papers or simulated test papers, suggestion book/), audio or audio-visual materials, or other supplementary teaching materials. By examining the materials chosen by the teachers, the present researcher tried to be aware of the contents of teaching.

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

175

In addition to the above activities and events listed in the observation schedule, other visible classroom events were recorded in the note-taking sheets (i.e., class notes). These were used for comparison with the characteristics of the HSC examination to determine whether the observed classroom phenomenon was related to the test. The observation participants disagreed to be audio and video recorded. All the observed lessons were recorded in writing. The observation instrument included observation schedules, note-taking sheets, pencils and a watch. During each observation, the observation schedule was filled in. The other raw and narrative data were also documented in writing.

4.2.2.2.3 Evaluation of Examination Related Documents

In this study, the present researcher conducted an intensive review of the examination related documents pertaining to the HSC syllabus and curriculum, HSC examination papers (English *First Paper* and *Second Paper*), 20 answer scripts, and the textbook *English for Today* for classes 11-12. The HSC examination-related documents, and the aims and objectives targeted by the NCTB are taken as official sources reflecting the EFL education intentions. One purpose of the review was to find out what the HSC examination set out to measure (e.g. linguistic knowledge or language use) and whether or not the HSC examination represented the curriculum. Another purpose was to identify the characteristics of the HSC examination, for they would serve as the basis for a comparison with what was happening in the classroom, and would help determine whether the observed classroom phenomenon was closely test-related (e.g., whether they were similar or there were gaps between the two).

4.2.2.2.4 In-depth Interview

Both watching and asking are very powerful instruments in any complex research such as washback study. In order to triangulate and possibly extend the findings of the present study, semi-structured interviews were conducted with 6 EFL teachers, 4 HSC examiners of EFL, and 3 curriculum specialists. They were all directly involved in HSC education in Bangladesh. This was an interview on a one-to-one basis. It was a supplementary instrument used in the research for eliciting

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

176

qualitative data on: how they planned, how they designed the policy, how they delivered inputs, and how they received outcomes. The different sets of semi-structured interview questions (for qualitative data) for EFL teachers (Appendix-4A), EFL examiners (Appendix-4B) and curriculum specialists (Appendix-4C) were designed; and the interviewees answered them in their own ways.

In qualitative research, "interviewing (i.e., the careful asking of relevant questions) is an important way for a researcher to check the accuracy of the impressions he or she has gained through observations" (Fraenkel & Wallen, 2003, p. 455). For the purpose of this study, as in other washback research (Watanabe, 2004), interviews allowed access to reasons behind some of the behaviours observed in the classroom during the research. The format of the interview followed the interview form used by Qi (2004) for her study of the washback effects of the National Matriculation English Test in China. This interview protocol was chosen as the model because of its construction. The researcher had a set of questions pertaining to how the testing programme might be affecting teaching; but in order to allow the participant's freedom of expression, and to avoid "leading" the participant with focused questions, the researcher engaged the participant in a dialogue, instead of a question and answer session. In the interview with the education planners, the researcher used a small set of questions.

In the qualitative paradigm, interviews provide opportunities for researchers to probe particular variables for detailed descriptions. Concerning the value of the data collected through interviews, Glesene and Peshkin (1992) argue that the potential strength lies in the fact that interviews provide opportunities to learn about the things that might be missed by the researchers to explore alternative explanations of what is seen. All of the interview sessions were noted down minutely in order to avoid missing the interviewees' comments. In this particular type of interviewing, the present researcher typically told the same questions to each of the participants used. Several reasons for using the structured interviews were:

1. The structured interviews are preferable when there is a limited period of time, and it is possible to conduct each interview only once (Patton, 1990).
2. The structured interviews are systematic (Patton, 1990; marshal and Rossman, 1989).

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

177

3. The structured interviews facilitate organization and data analysis as the format of the interview allows researchers to locate each format's response to the same question quickly (Patton, 1990).

4. The standardized interviews increase comparability of responses as each informant is asked the same question (Patton, 1990).

The assent was obtained from all of the participants before the interviews took place. The researcher himself was the moderator and took detailed notes throughout the discussion, including notes on the participants' body language. All of the interview sessions were noted down minutely in order to avoid missing the interviewees' comments.

4.3 Pilot Study

The present researcher conducted pilot study, and used the test-retest method to compute the reliability of the survey instrument. The initial versions of the questionnaires were first piloted in March 2010 on 20 students and 10 higher secondary-level EFL teachers to check the appropriateness of the questions. The results of the pilot study indicated that they were suitable to administer. Yet, some of the student respondents opined that they could not understand the message of 2/3 questions; therefore, they took help from the researcher to understand them. Based on the information gained from the pilot study, they were refined, reworded, revised and reframed for clear understanding; and were administrated once again on the same group three weeks later. Spearman's (1947) coefficient of correlation formula was used in order to find out the reliability coefficient, and the ratings were considered to be sufficient for the purpose of applying the questionnaires.

4.4 Ethical Considerations

Ethnical issues involved in collecting data, conducting research, and reporting the results were taken into careful consideration. The selection of participants was largely based on their willingness and interest to share their class activities with the present researcher. Early in the interviews, the present researcher informed all the potential participants of the purposes of the research and also

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

178

informed them and their respective schools that their identity would be kept concealed through use of pseudonyms. Assurance was given that the confidentiality of each participant's intellectual property and privacy would be maintained throughout the study. The curriculum specialists conditioned that their name should not be disclosed and mentioned in this thesis. The participants' name, identity and their comments were handled with due importance and care.

4.5 Timeline and Data Collection Procedures

The data for the present study was collected under a planned procedure and schedule. All the data was collected during February 2010 to November 2010. The researcher used a number of instruments (e.g. questionnaires, classroom observation, in-depth interview, and review of the HSC examination related authentic documents) and collected data from a number of sources (e.g. students, teachers, EFL examiners, curriculum specialists, question papers, answer scripts, textbook, syllabus and curriculum). The analysis of test related authentic documents, the pilot study, the questionnaire survey, classroom observation were all interdependent and interrelated for the study. The research sites were designed both in urban and rural areas. The table below (Table 4.5) shows how quantitative and qualitative data were collected at different stages throughout the data collection process:

Table 4.5: The data collection procedures

Data Collection Phases	Activities/ Procedures	Timetable
Phase-1	• Mapping of the site and sample selection • Baseline data • Review of examination related documents (question papers, answer scripts, textbook, syllabus and curriculum, etc) • *Literature review* • *Pilot study* • Planning for survey administration	**January 2010-March 2010**

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

179

Phase -2	• Visiting survey sites • Seeking permission from authority • *Questionnaire Survey Administration* • Planning for classroom observation • Adopting, drafting and finalizing the Observation Schedule (COLT, Part- A; Modified UCOS; and Semi-structured Checklists)	**April 2010– July 2010**
Phase- 3	• *Classroom Observations in 10 sites* • Planning for conducting interviews • Drafting semi structure questions for interviews	**August 2010- September 2005**

Phase -4	1. *In-depth interview*
	   • EFL teachers
	   • EFL examiners
	   • Curriculum specialists
	2. *Data analysis methods and procedures framed*
	3. *Data from review of documents, and part of questionnaire survey data analysed*
		October 2010- November 2010

This research design is principally a sequentially exploratory triangulation design (Creswell & Plano-Clark, 2007). It is sequential because phases of data collection follow each other in a specific sequence over time. When regarded horizontally at each stage, this design has concurrent elements, with quantitative and qualitative data collection. Table 4.5 provides a specific timeline of when each data collection period took place. It also lists the various sources of data, both quantitative and qualitative, that were obtained throughout the entire data collection process. The data collection and data analysis procedures are explained in greater detail in the following sections:

Phase -1: The first stage of data collection involved the review and analysis of EFL testing and teaching related documents at the HSC level. The washback effect of the HSC examination at the macro level (e.g., current social and educational context) was examined. The goal of this stage of data collection was to get a broad and holistic understanding of washback and its influence on the EFL teaching–learning areas, and objectives of the syllabus and curriculum, textbook materials, lesson contents, characteristics of the HSC question papers in English, etc. At this stage, the researcher also obtained baseline data from different sources to provide a comparison with the data to be collected later for assessing examination

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

180

washback. The researcher carried out a pilot study during this stage. The pilot study took place in February 2010 to March 2010. The first pilot study was conducted during 13 February to 21 February 2010, and again three weeks later during 10 March to 16 March, 2010.

Phase- II: The second stage involved the administration of a questionnaire survey. The survey was carried out in different higher secondary colleges through two questionnaires. The researcher visited 18 colleges in urban and rural areas, and collected data from the higher secondary students and teachers. The present researcher distributed typed questionnaires to the respondents, and requested to provide information spontaneously. Survey data collection took place during April 2010 to July 2010. All the questionnaires were administered in the face-to-face classes. Data collection took place without any interference of teacher or the researcher, and thus the researcher guaranteed the reliability of the results. While administering questionnaire survey in different sites, the researcher was planning to conduct the classroom observation. At this stage, he finalised the observation schedules and checklists, and selected the 10 research sites of which 5 sites were in rural areas and 5 were in urban colleges. When the data was collected, the scripts were processed for analysis and interpretation.

Phase- III: At the third stage, the classroom observations were conducted in the selected sites. The washback effect at the micro level (e.g., the impact of the HSC examination on classroom teaching and learning) was investigated. Immediately after the classroom observations, in-depth interviews were conducted with the six observed teachers. The data derived from the first and the second stages were taken as the baseline data for this study, and they would be compared with how teachers taught after they had responded to the questionnaire. During this phase (Phase-III), the present researcher conducted a broad spectrum of observations, and chose a true representative sample. At this stage, the focus of the study evolved from an initially broad and holistic set of ideas to more specific questions related to the teachers' reactions to the examination in English. This round of observations was conducted in August 2010 to September 2010. During this phase, he also planned to conduct the in-depth interviews, and drafted the best-suited questions to be asked during the interviews.

Phase-IV: The last stage of data collection consisted of in-depth interviews. The purpose of this stage of data collection was to confirm the salient and recurring themes and patterns that had emerged from the data gathered in the earlier stages and to see if the teaching of the target test features accelerated right before the test. In this stage, all data sources were cross-examined to finally develop a theory to explain the findings. The researcher conducted the interviews during the October 2010 to November 2010. During this stage the data analysis methods and procedures were finalised and framed. In this stage, the data collected from the review of documents and questionnaires survey were analysed. When the data were collected, the scripts and raw data were processed for analysis and interpretation.

4.6 Data Analysis

A mixed methods (MM) approach combining the qualitative and quantitative methods was used both for data collection and data analysis in this study. According to Bogdan and Biken (1998), data analysis is the process of bringing order, structure, and meaning to the mass of collected data. This process entails uncovering patterns, themes, and categories. The review of the literature has demonstrated that there are multiple facets of change of washback that occur at the systemic level as well as within the school and classroom contexts. It was felt that this methodology would be the best suited for capturing the complexity of the processes inherent in educational change. Firstly, a close examination of the pertinent documents (sample test of the HSC examination, NCTB formulated curriculum, textbook, past examination questions, answer scripts, etc.) was performed. An intensive analysis of the characteristics of the HSC examination in English was made; and is reported in the next chapter.

Secondly, qualitative analyses of the classroom observation data as well as the in-depth interview data were conducted. The analyses involved the use of the constant comparative method (Bogdan & Biklen, 1998; Glasser and Strauss, 1990; Lincoln & Guba, 1985) in which the data were classified into categories. Specifically, the researcher used inductive logic to identify and categorize emerging themes, perspectives and events from a mass of narrative data. Thirdly, quantitative analyses were performed, which involved frequency counts (and/or percentages by

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

182

category), descriptive statistics and the following inferential statistical procedures: Levene's Test for Equality of Variances, and T-Test. These were applicable to this study because they were commonly used to analyse interrelationships among large numbers of variables and to explain these variables in terms of their common underlying dimensions.

The science of statistics assists researchers in planning, analyzing, and interpreting the results of their investigations. It provides accurate information about the problem that arouses one's interest. The investigator collects and analyses the data applying appropriate statistical procedures. In the present study, the data were analysed using the SPSS 18.0 for Windows; the descriptive statistics were also used to analyse the responses of the participants. Data were analysed in two phases. Qualitative analysis involved the use of a constant comparative method, while quantitative analysis in this study involved descriptive statistics (e.g., frequency counts, means, standard deviations, etc.). After this initial step, the responses of the participants for each statement were tabulated and converted into percentages. The percentages were then tabulated and graphed to allow a clear view and understanding at a glance of how the responses were distributed across the two groups of participants- teachers and students. Since the responses were actually on a binary scale, the two categories of agreement (Strongly Agree &Agree) and disagreement (Strongly Disagree & Disagree) were respectively collapsed to allow for easier discussion of the results.

Finally, the different types of data sources were synthesized and integrated. To be specific, the qualitative data (through interviews and observations) were compared with the quantitative data (through the questionnaires) in search of patterns of agreement and disagreement. The purpose of the comparison was to find out whether the results from the qualitative data analysis were congruent with those from the quantitative data analysis. As a result of the comparison, the categories were combined and reorganized based on the common features found. The results of the comparison were presented with visual aids (charts, tables, etc.). The data were reviewed in a timely manner so that they could inform subsequent stages of the data collection process. More details of how the data were analysed are reported below in Table 4.6:

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

183

Table 4.6: Data analysis procedure

Analysis of the Document	Case Studies	Questionnaire Five-point Likert scale	Integration of Data
(e.g. the HSC Syllabus and Curriculum, textbook, HSC exam papers in English) ●Goals ●Contents ●Skills ●Methodology	Observation ● Coding ● Frequency counts	Closed items (Likert Scale) ● SPSS 18.0 used ●Descriptive stats (frequency counts, SD) ●Inferential stats (Levene's Test, T-Test)	Questionnaire + Interview + Observation Qualitative and quantitative
In-depth Interview • Organizing data • Categorization • Developing theory		Interview Question (Open-ended questions)-Constant comparative method • EFL teachers • EFL examiners • Curriculum specialists	

4.6.1. Analysis of Questionnaire Data

The questionnaire survey data were analysed in multiple ways. Descriptive statistics, inferential statistics, tables, charts, and graphs were applied to clarify and explain the analysis. Survey results can be presented in different ways: by text, in tables, in figures in charts, graphs and histograms. Tables and figures are useful methods to convey data when the reader or viewer is required to take in information while reading or listening. The tables and graphs can describe larger sets of numbers better than text, and should be used if trying to communicate more than three or four numbers. The computer program Statistical Package of the Social Sciences (SPSS, 18.0) for Windows was used to compute descriptive statistics and perform inferential statistics. A detailed discussion of all these procedures is provided in Chapter Five.

4.6.1.1 Descriptive Statistics

When dealing with the questionnaire data involving various components (such as the syllabus and curriculum, teaching methods, textbook materials, beliefs of test impact on teaching/learning and pedagogical knowledge, etc.), the present researcher first relied on frequency counts to know about the frequencies and

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

184

percentages of the teachers' and the students' responses by category, and also examined the mean and standard deviation (STDV) of each question.

4.6.1.2 Inferential Statistics

The Levene's test for equality of variances and T-Test for equality of means (independent samples test) were performed to examine whether the means of two groups (teachers and students) responded to questions. The independent samples T-Test compare the mean scores of two groups on a given variable. For the independent samples T-Test, it is assumed that both samples come from normally distributed samples with equal standard deviations (or variances). A normally distributed variable is assumed to have a skewness and kurtosis near zero (Arbuckle, 2006). Reliability for internal consistency was calculated using the Cronbach's (1970) Alpha Coefficient.

4.6.2 Analysis of the Data from Classroom Observations

The data from the classroom observations were first coded according to the categories developed in the observation schedule. Then, frequency counts were applied based on these labelled categories. The analysis involved a calculation of the duration of each classroom activity and instructional pattern in an average percentage of the class time. After that, the percentages of the time spent on each of the categories on the observation schedule were compared to determine the frequency of occurrence of various classroom interaction patterns and activities. After this analysis, the observation data were compared to the data derived from the interviews to see whether they were compatible to each other. As Maxwell (1996) indicated, compatibility of interviews or observations is important.

4.6.2.1 Analysis of Data from COLT, UCOS, and Checklists

Data collected with the COLT and UCOS observation schedules were processed in different components. Besides, the data collected through the self-prepared observation checklists were compared whether they were overlapped or gone beyond systemic analysis. The instrument COLT (Part-A) was completed with

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

185

the observer coding the classroom events as they occurred. Detailed notes of the activities and episodes were taken during the lessons mainly focusing on the communicative features of classrooms. The points of observations were placed in separate categories.

When deciding on the coding of data according to coding categories, it was necessary to reduce the categories in a standardised way. Additional notes taken during the observation and materials collected from the classes were used to inform decisions when identification of an instance was not clear simply from the basic field notes alone. The data collected at this stage was somewhat qualitative, and as such the process was an iterative one. Classroom observation data was recorded in rows and columns in Excel files. Due to the varying lengths of the classes and courses, all the activities were expressed as the percentage of the overall class time. Once the data had been analysed quantitatively, a brief summary of the course was written. The details of the analysis are presented in Chapter Five.

4.6.3 Analysis of the Data of Examination Related Documents

The analysis of the examination related documents aimed at identifying the characteristics of the documents and their relations to classroom teaching, learning and the HSC examination in English, for they would serve as the basis for a comparison with what was happening in the classroom, and would help determine whether the observed classroom phenomenon was closely test-related. The researcher applied different criteria, checklists and guidelines (Appendices 3A to 3F) to review the examination related documents. The analyses of examination related documents determined how these materials influenced the academic behaviours of the teachers and the learners, and exerted washback on EFL teaching and learning at the HSC level.

4.6.3.1 Analysis of the Syllabus and Curriculum

The present researcher analysed the HSC English syllabus and curriculum following some set guidelines (Appendix -3A) posed by a number of researchers (e.g. Porter, 2002, 2004; Richards, 2001; Brown 1995; 2007). A syllabus refers to the content or subject matter of an individual subject, whereas a curriculum refers to

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

186

the totality of contents to be taught and aims to be realised within one school or educational system. A syllabus is a specification of the content of a course of instruction and lists what will be taught and tested. In Bangladesh, the HSC English syllabus directly corresponds and represents to the HSC English curriculum- hence the HSC English syllabus and curriculum can be used interchangeably. So, they are both used as a mutual term in this research. The term "curriculum" in this study is seen to include the entire teaching/learning process, including materials, equipment, examinations, and the training of teachers.

Porter (2004) defines curriculum analysis as the systematic process of isolating and analysing targeted features of a curriculum. Any curriculum analysis most commonly involves describing and isolating a particular set of contents (e. g. language arts content) in a curriculum and then analysing the performance expectations, or cognitive demand, that describe what students are to know and do with the content. Content, is defined as the domain specific declarative, procedural, tactile and situative knowledge targeted by a curriculum. Performance expectations are generally defined as the level at which a student is expected to know and employ the content as a result of the instructional activities and assessments conducted in the curriculum. Through systematic analysis of curricula, educators can begin to compare and contrast various aspects across multiple curricula. Porter (2002, 2004) also makes distinctions regarding the four levels at which curricula analysis may occur. The four levels at which one may analyse a curriculum include intended, enacted, assessed, and learned. The method introduced in this study was only concerned with examining the intended curriculum.

Curriculum and syllabus analysis is a type of methodology within qualitative research. The present researcher followed a systematic process for completing a language-based curriculum analysis to address a critical review of the curriculum expectations which might challenge students with communication difficulties. This analysis leads to the development of strategies for making modifications in the presentation of curriculum material. The history of curriculum development in language teaching starts with the notion of syllabus design. A syllabus design is one aspect of curriculum development but is not identical with it. The present researcher used the following steps to analyse the HSC EFL curriculum and syllabus:

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

187

Needs Analysis

The researcher conducted "needs analysis" of the HSC English syllabus and curriculum because it was a fundamental point to be analysed. Richards (2001) suggests "Needs Analysis" is fundamental to the planning of general language courses. In language curriculum development, Needs Analysis serves the purposes of (i) providing a mechanism for obtaining a wider range of input into the content, design and implementation of a language program through involving such people as learners, teachers, administrators and employers in the planning process, (ii) identifying general or specific language needs which can be addressed in developing goals, objectives, and content, for a language program, and (iii) providing data which can serve as the basis for reviewing and evaluating an existing programme.

Goals Setting or Objectives

The second step in the curriculum analysis process is to establish goals or objectives. The present study examined the goals and objectives of the HSC English curriculum to evaluate its standard with regard to communicative language teaching and testing. According to Brown (1995, p. 71) goals are broader in their concept as they are general statements concerning desirable and attainable programme purposes and aims". Objectives on the other hand are much more specific than goals, both in their conception and in their context. Objectives usually refer to aims and purposes within the narrow context of a lesson or an activity within a lesson. Furthermore Graves, (2000, p.93) adds that the goals and objectives are "not set in cement" but "should be clearly stated, as teachers hope to accomplish given what they know about their context, about "students' needs and our beliefs about how people learn", and finally "our experience with the particular content".

Content and Methodology

A curriculum advocates teaching methods to be used in the class. The teaching methods are recommended on the basis of contents, and the goals and objectives of the syllabus and curriculum. The present study examined which teaching method was recommended to achieve the goals and objectives of the syllabus and curriculum. The study also analysed the contents to be taught in the class. Richards (2001) points out that there are two major forms of curriculum models and teaching method. In the Educational Curriculum context "Methodology

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

188

is concerned with choosing learning experiences, activities and tasks, which lead to mastery of the linguistic content of the syllabus, and at the same time, attain the objectives of the language program" (Richards, 2001, p. 15).

Assessment/Testing

Assessment is an essential part of any curriculum. The present researcher reviewed how the present HSC English curriculum treated EFL testing. Brown (2007) argues "no curriculum should be considered complete without some form of programme evaluation". He adds that there are three interdependent elements to assess: students, teachers, and programme. Each of these relies on the both the others to be successful, or, conversely contribute to their failure.

Brown (2007, p. 159) explains that there are three possible ways that need to be considered in evaluating the success of the curriculum. First, everybody needs to be consulted ("all the stakeholders/participants"). Secondly the researcher needs to consider the "audience of the evaluation". Finally, the researcher needs to consider various aspects (Brown, 2007) of the programme evaluation as the following: appropriateness of the course goals, adequacy of the syllabus to meet those goals, textbooks and materials used to support the curriculum, classroom methodology, activities, procedures, the teacher's training, background, and expertise, appropriate orientation of teachers and students before the course, the students' motivation and attitudes, the students' perceptions of the course, the students' actual performance as measured by assessments, means for monitoring students' progress through assessments, institutional support, including resources, classrooms, and environment, and staff collaboration and development before and during the course.

4.6.3.2 Analysis of *English for Today* for Classes 11-12

The present study analysed *English for Today* for classes 11-12 to look into whether the textbook corresponded to the HSC English syllabus and curriculum. The analysis also tried to find out if the HSC examination adequately communicated the lesson objectives of the English textbook. For the textbook analysis, a checklist (Appendix-3B) was applied which was adapted from the American Council on the Teaching of Foreign Languages (ACTFL). A number of textbook evaluation checklists and guidelines had also been studied to evaluate *English for Today* for

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

189

classes 11-12. Bailey (1999) advocates that textbook washback is a possible result of test use. She suggests that test preparation materials are the indirect evidence of washback. The textbook should give introductory guidance on the presentation of language items and skills. It serves as a syllabus. The analysis looked into whether HSC examination in English had any washback (positive or negative) on *English for Today* for classes 11-12.

4.6.3.3 Analysis of the HSC English Test

The framework proposed by Bachman and Palmer (1996) is often taken as a theoretically grounded guideline (Appendix-3E) for analysing the characteristics of a test. This conceptual framework consists of a set of principles involving five facets of tasks: setting, test rubric, input, expected response, and relationship between input and response. But here, the present researcher presented and discussed four features in particular which he thought crucial for this study. A test is a part of curriculum, so, the test should reflect and correspond to the syllabus and curriculum. The present study performed the HSC English test: *First Paper* (Appendix-3C) and *Second Paper* (Appendix-3D) analyses to examine the nature, contents, characteristics, and their influence (washback) on classroom teaching and learning.

4.6.3.4 Analysis of the HSC Answer Scripts

The present researcher conducted "Answer Scripts" analysis to examine whether the examiners' evaluation/scoring system influenced teaching and learning. The researchers analysed 20 answer scripts of English *First* and *Second Paper* examined by 4 EFL examiners. The present researcher also observed the scoring/ marking procedures of the examiners. Afterward, the examiners were interviewed through semi-structured questionnaire.

Answer script analysis offers in-depth knowledge of the student as a learner on a prescribed course. It can include evidence of specific skills and other items at one particular time and language performance and progress over time, under different conditions, in all four modalities (such as reading, writing, listening, and speaking) or all three communication modes (interpersonal, interpretive, and presentational). Cheng (2004) suggests that analysis of answer sheets/scripts reflects

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

190

students' overall achievement in second or foreign language learning. Like classroom observation, answer sheet analysis is of great value. Bailey (1999) points out that answer sheet analysis is closely linked to instruction, which has two educational benefits. First, linking assessment to instruction means that what is being measured has been taught. Second, it reveals any weaknesses in instructional practices. Andrew (2004) suggests answer per analysis promotes positive student involvement. It is actively involved in and reflecting on their own learning. Li (2009) suggest that answer paper focuses how much positive or negative washback dominates the classroom activities.

Brown (2000) opines that answer papers are the visible evidence of learners' learning outcome. Enright (2004) suggests answer pages highlight how much communicative competence has been achieved opposed to how much it is tested. However, Morrow (1991) argues that answers to tests are more than simply right or wrong, and that they should be assessed on the basis of how far toward an approximation of the native speaker's system they have moved. Tests should reveal the quality of the testee's language performance. For the answer scripts analysis, a checklist (Appendix- 3F) was applied. The checklist was adopted in accordance with the guidelines of Morrow (1991) and Brown (2003).

4.6.4 Analysis of the Data from Interviews

In general, the data derived from the interviews (e.g. individual) as well as the data from classroom observations were analysed qualitatively by searching for themes and patterns. In the meantime, they were reduced and synthesized using focused summaries pertaining to the research questions and other emerging issues. The general aim of conducting interviews was to explore the breadth and range of views represented by the participants on the topic of the complexity of washback phenomena in relation to the HSC examination and English language teaching and learning. The interviews were also used for the collection of straightforward factual information. Oral consent was obtained from all participants prior to interviews. Face-to-face interviews were then conducted with two EFL teachers, examiners and curriculum specialists. Those participants were members of the target population but not part of the final sample in the main study. The purpose of interviews with teachers was to explore the teachers' beliefs: whether teachers believed that their

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

191

teaching had been influenced by the HSC examination in English. The interviews also provided an opportunity for the teachers to give their impressions of the lessons, to describe the rationale behind their choices of activities and materials, and to express their opinions regarding the imposition of English tests as a graduation requirement. The copies of the interview schedule are given in Appendix section (4A, 4B, 4C).

4.6.4.1 Design and Procedure of the Interviews Analysis

All of the interview questions were derived from the review of the literature and contacts at the preliminary information-gathering stage, and there were parallels between questions in the questionnaires and interviews. All of the interviews were semi-structured with prompts whenever necessary and they were conducted in English and Bengali, and hence the language in which all participants would most likely feel comfortable communicating. All the interviews were audio-recorded and backed up by written field notes in order to trial the data collection procedure and the equipment. At this pilot stage, interviewees expressed no particular difficulties in answering any of the questions. Therefore, the interview schedules were employed for the main study with just occasional minor corrections of wording. All of the interviews lasted about 20 to 30 minutes. Each participant was interviewed once and the interviews were audio-recorded.

All of the interviews of study were transcribed in full in the original language and then translated into English by the researcher. As suggested by Gillham (2005), the transcripts were edited by avoiding repetitions and putting substantive statements in chronological order to make grammatical sense, which facilitated further levels of analysis and provided a relatively tidy and accessible form for interpretation. Morse and Richards (2002) distinguished between three kinds of coding: descriptive coding, topic coding and analytic coding. The process of analysis began with topic coding. The topics were designated according to the categories previously used in designing the interview schedules. The categories were used as preliminary ways of understanding the data as "at the beginning of a study the researcher is uncertain about what will ultimately be meaningful" (Merriam, 1998, p. 179). The researcher then looked for patterns across each of the categories, seeking to identify recurrent

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

192

analytical categories. The transcripts were then grouped and edited again according to the new analytic categories.

For the purpose of examining the reliability of the interview data, the researcher went back to the audio-recorded interviews and recoded the previously analysed interviews. The purpose of this approach was to make sure that the present researcher had been consistent with the criteria for analysis. The main study interview data are presented and discussed in Chapter Five (section 5.4). The qualitative data analysis proceeded along the following steps:

4.6.4.1.1 Organizing the Data

First, the researcher performed minor editing to make field notes and interview summaries manageable and retrievable. Then, he closely examined a small batch of data, and jotted down the emerging themes and patterns. Having developed some preliminary categories of themes, he read through the data, and grouped them according to these categories. He analysed the data logically, and assigned units of data into categories based on shared themes. The method that he used to analyse the data is called the constant comparative method (Strauss & Corbin, 1990). The remarks and assertions made by interviewed personnel/examiners during the various interview sessions were constantly compared and contrasted throughout the research process.

4.6.4.1.2 Developing Theories and Reporting the Outcomes

This step involved simplifying the codes and reducing the number of categories. Specifically, smaller categories were merged into a larger category. This procedure of combining and recombining the categories entailed data reduction. Eventually, this systematic process of induction enabled the present researcher to relate the data to a theory. Drawing on the coding system developed by Strauss and Corbin (1998), he was able to build theoretical explanations, develop concepts and propositions from data. As a result, a grounded theory was developed at this stage. It provides a thick description of the research settings and a comprehensive account of the results. A holistic perspective was adopted when it came to presenting the participants' perspectives and views.

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

193

4.7 Conclusion

Research outcomes largely depend on the methodology a study applies. Methodology differs from subject to subject and context to context. Since the context may have an impact on results, the researcher needs to be informed of what measures what. Therefore, the present researcher was very careful in applying an appropriate methodology for this research. His attempt was to ensure that methods and approaches utilised were appropriate to capture the washback traces. In general, there are two types of designs adopted by washback researchers: mono-method and mixed methods approaches.

It is a fact that there were not many existing instruments in the area of washback which could be drawn upon. No single uniform questionnaire has emerged as being widely used to survey either teachers or students about language testing washback. Bailey (1999) pointed out that it would be a valuable contribution to the available methodological instruments for washback study to develop a widely usable questionnaire for teachers and for students.   The subjects of the present study, the method, the instrument, data analysis procedures, are all validated and supported by the previous research studies carried out during the last decade.

This chapter has presented and discussed several aspects of the research design adopted in the present study. First, an introduction is given to the application of a mixed-methods and emergent design. It has indicated that a mixed-methods strategy is appropriate to this study since each single method has its individual weaknesses. Second, some general background information is given about the participating students, teachers, other professionals, and research sites. Third, a description of the instruments is given, along with a brief rationale for using them. Fourth, the procedures of data collection are explained. The final section has provided a description of the procedures and methods of data analysis.

After analysing all types of data, the researcher made a comparison among data from different methods in order to triangulate and complement the findings. If findings were congruent, interpretations could be made on the basis of the consistent results. When the data showed inconsistency, the researcher tried to speculate on the underlying reasons, and interpreted the divergent results. The next chapter presents and discusses the findings of the study.

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

194

Chapter Five

Presentation and Discussion of the Findings

The methods applied to collecting data in the present study have been detailed in the previous chapter. This chapter presents and discusses the findings of the analysis of data collected from varied population and sources in separate sections. It begins with the presentation and discussion of the quantitative findings derived from the questionnaire surveys. After that, the qualitative findings resulted from the classroom observations, analysis of examination related documents and interviews with teachers, examiners, and curriculum specialists are presented and discussed. Given the substantial amount of data yielded from this study, a detailed description of all of the findings of this research is beyond the scope of this thesis. The present researcher was compelled to limit the presentation of results in this thesis to only the findings that specifically addressed the research questions.

5.1 The Questionnaire Surveys

As introduced in Chapter Four, a survey was administered to the participating students and teachers in this study to poll their beliefs in the HSC examination in English, and their opinion of its influence on EFL education, their views of language teaching and learning, and information about what they considered to be effective ways of teaching. Five hundred students and one hundred twenty five teachers took part in the survey. Both the teachers and the students responded to the questionnaires related to the syllabus and curriculum, materials, teaching methods, teaching methods, classroom tasks and activities, language skills and element, and respondents' beliefs, attitudes and perception as to the test. This section presents the results of statistical analyses. For the purposes of reporting, the decimal numbers calculated were rounded off to the nearest whole number. In the present study, the internal consistency was measured based on the correlations between different items of the student and teacher questionnaires. The questionnaires comprised 6 sections on 6 domains. Under each section, there were several questions. The internal consistency of every section was measured statistically.

195

Internal consistency reliability defines the consistency of the results delivered in a test, ensuring that the various items measuring the different constructs deliver consistent scores. Internal consistency reliability is a measure of how well a test addresses different constructs and delivers reliable scores. In this study, the internal consistency has been measured with Cronbach's alpha, a statistic calculated from the pair-wise correlations between items. Internal consistency ranges between zero and one.

A commonly accepted rule of thumb is that an α of 0.60-0.70 indicates acceptable reliability, and 0.80 or higher indicates good reliability. High reliabilities (0.95 or higher) are not necessarily desirable, as this indicates that the items may be entirely redundant. The items produced a reliability estimate of 0.74 (textbook materials) to 0.90 (EFL skills and elements) for teachers, above the desirable threshold of 0.70 (Garson, 2007). The student items reliability ranged from 0.62 (Teaching methods and approaches) to 0.88 (EFL skills and elements). The magnitude of the relationship investigated in the study was described on the basis of the scale delineated by Davies (1971) as shown below:

1.0-------------------- Perfect
0.70---------------0.99 Very high relationship
0.50---------------0.69 Substantial relationship
0.30---------------0.49 Moderate relationship
0.10---------------0.29 Low association
0.01---------------0.09 Negligible relationship

The present study used two questionnaires: student questionnaire and teacher questionnaire. Every questionnaire had six sections comprising altogether 45 questions. The internal reliability of the questions of every section is as follows:

Table 5.1: Reliabilities estimates

SL	Sections	Items	Students	Teachers
1	Syllabus and Curriculum	7 items	0.79	0.78
2	Textbook Materials	10 items	0.76	0.74
3	Teaching Methods and Approaches	8 items	0.62	0.70
4	Classroom Tasks and Activities	6 items	0.85	0.87
5	EFL Skills and Elements	5 items	0.88	0.90
6	Students' Belief, Attitudes and Perception as to the test	8 items	0.71	0.77

5.1.1 The Statistical Analysis

The findings of the study are presented as per themes. The quantitative analysis in this study involved descriptive statistics (e.g., frequency counts, means, standard deviations, etc.) and inferential statistics. The SPSS 18.0 for Windows was used for the statistical analysis.

The responses of the participants for each statement were tabulated and converted into percentages. The percentages were then tabulated and graphed to allow a clear view and understanding at a glance of how the responses were distributed across the two groups of participants. Since the responses were actually on a binary scale, the two categories of 'strongly agree' and 'agree' were collapsed into single category *agreement*, while 'strongly disagree' and 'disagree' were collapsed into single category *disagreement* to allow easier discussion of the results. The statements assessing the expected response of the participants were adopted through a five-point Likert scale (Likert, 1932). On the scale, statements were coded as Strongly Agree=5, Agree=4, Neutral=3, Disagree=2, and Strongly Disagree=1. Five experts (the supervisor, two senior researchers, and two statisticians) in statistics were consulted in identifying the analytical levels of estimating values of mean scores of each item in the instrument (i.e. questionnaire). What needs to be mentioned here is that the questionnaire statements are reported as if they were questions. For instance, Q1 refers to Statement Number 1.

In the study, the present researcher performed analyses of different issues such as teachers' beliefs, knowledge and experience. However, due to the limited scope of this paper, they are not presented in this thesis. Here, only six major themes are reported: (1) the syllabus and the curriculum and its relation with the HSC examination in English and its (examination) impact; (2) textbook materials and washback effects of the HSC examination on their teaching and learning; (3) teaching methods, respondents' beliefs in teaching and learning, and the ways they teach; (4) classroom activities and knowledge base; (5) practices of language skills and elements; and (6) respondents' belief, attitude, and perception towards test. All the themes pertained to the research questions that were posed in this study.

Some relevant statistical tests had been conducted for data analysis and to draw reliable findings from the current research. Mean (M) scores, mode, median,

standard deviation (STDV), variance, skewness, kurtosis, etc. were mainly performed for the analyses of the data. Some inferential analyses such as reliability, correlation coefficient, Levene's Test for Equality of Variances, T-Test significance were performed in the study.

For every question, the mean score was calculated to support the frequency of the findings. The mean score is the average and is computed as the sum of all the observed outcomes from the sample divided by the total number of events. The mean (M) is a *weighted average*, with the relative frequencies as the weight factors. A distribution can be compared with a mass distribution, by thinking of the test marks as point masses on a wire (the *x*-axis) and the relative frequencies as the masses of these points. In this analogy, the mean is literally the centre of mass--the balance point of the wire. Usually, x is used as the symbol for the sample mean. With this in mind, it is natural to define the mean of a frequency distribution by-

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x$$

In statistics, 'n' is the sample size and the x corresponds to the observed value.

The study calculated the variance and the standard deviation. Both are measures of the spread of the distribution of the mean. The physical unit of the variance is the square of the physical unit of the data. The researcher calculated standard deviation (STDV) because it was a widely used measurement of variability or diversity used in statistics and probability theory. It shows how much variation or 'dispersion' there is from the average (mean or expected value). Standard Deviation (STDV) is the extent to which data differ from the mean.

It should be mentioned that a low standard deviation indicates that the data points tend to be very close to the mean, whereas a high standard deviation indicates that the data are spread out over a large range of values. Standard deviation measures spread in the same physical unit as the original data both measures of spread are considered very useful for the study. The variance is defined to be-

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x - \bar{x})^2$$

and, the standard deviation is defined to be-

$$s - \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x-\bar{x})^2}$$

The standard deviation is a measure of how the data is clustered of the mean. For large sets of data, approximately 68.3% of the data lies within one standard deviation of the mean and approximately 95.4% of the data lies within two standard deviations of the mean.

The fundamental task in the statistical analyses for the present study was to characterise the location and variability of a data set. A further characterization of the data includes skewness and kurtosis. Skewness is a measure of symmetry, or more precisely, the lack of symmetry. A distribution, or data set, is symmetric if it looks the same to the left and right of the central point. Kurtosis is a measure of whether the data are peaked or flat relative to a normal distribution. That is, data sets with high kurtosis tend to have a distinct peak near the mean, decline rather rapidly, and have heavy tails. On the other hand, data sets with low kurtosis tend to have a flat top near the mean rather than a sharp peak. A uniform distribution would be the extreme case. The histogram is an effective graphical technique for showing both the skewness and kurtosis of data set. For univariate data Y_1, Y_2,..., Y_N, the formula for skewness is:

$$skewness = \frac{\sum_{i=1}^{N}(Y_i - \bar{Y})^3}{(N-1)s^3}$$

Where $\bar{Y}$ is the mean, s is the standard deviation (STDV), and N is the number of data points. The skewness for a normal distribution is zero, and any symmetric data should have skewness near zero. Negative values for the skewness indicate data that are skewed left and positive values for the skewness indicate data that are skewed right. By the data skewed left, we mean that the left tail is long relative to the right tail. Similarly, the data skewed right means the right tail is long relative to the left tail. Some measurements have a lower bound and are skewed right. For example, in reliability studies, failure times cannot be negative.

Kurtosis characterizes the relative peakedness or flatness of a distribution compared with the normal distribution. Positive kurtosis indicates a relatively peaked distribution. Negative (-) kurtosis indicates higher kurtosis means more of the variance is due to infrequent extreme deviations, as opposed to frequent

modestly-sized deviations. A high kurtosis distribution has a sharper "peak" and fatter "tails", while a low kurtosis distribution has a more rounded peak with wider "shouldersa" relatively flat distribution.

$$\text{Kurtosis} = \frac{\sum (x_i - \bar{x})^4}{ns^4}$$

Here, $\bar{Y}$ is the mean, s is the standard deviation, and N is the number of data points.

Skewness characterizes the degree of asymmetry of a distribution around its mean. Positive skewness indicates a distribution with an asymmetric tail extending towards more positive values. In this example, the researcher compared several well-known distributions from different parametric families. Negative skewness indicates a distribution with an asymmetric tail extending towards more negative values. The skewness statistic is sometimes also called the skewedness statistic. As the skewness statistic departs further from zero, a positive value indicates the possibility of a positively skewed distribution (that is, with scores bunched up on the low end of the score scale) or a negative value indicates the possibility of a negatively skewed distribution (that is, with scores bunched up on the high end of the scale).

If skewness is positive, the data are positively skewed or skewed right, meaning that the right tail of the distribution is longer than the left. If skewness is negative, the data are negatively skewed or skewed left, meaning that the left tail is longer. If skewness is zero (= 0), the data are perfectly symmetrical. But a skewness of exactly zero is quite unlikely for real-world data. Bulmer, M. G., *Principles of Statistics* (Dover, 1979) suggests this rule of thumb:

a.) If skewness is less than −1 or greater than +1, the distribution is highly skewed.

b.) If skewness is between −1 and −½ or between +½ and +1, the distribution is moderately skewed.

c.) If skewness is between −½ and +½, the distribution is approximately symmetric.

d.) With a skewness of −0.1098, the sample data for student heights are approximately symmetric.

The present researcher also presents the findings from the inferential statistical analyses. Levene's Test for Equality of Variances and Independent Sample Tests (T-Test) were performed for some advanced level of analysis of data.

5.1.2 The Syllabus and Curriculum

The findings of the syllabus and curriculum have been discussed and analysed in this section. The findings have been tried to be validated through cross referencing. Through the interpretation of the findings, the nature and scope of washback of the HSC examination on EFL teaching and learning in general and on the syllabus and curriculum in particular have been examined.

The test always follows and does not lead the curriculum (Lindvall and Nitko, 1975). Given an inappropriate test, narrowing of the curriculum impedes teaching and learning EFL/ESL (Smith, 1991). Since test contents can have a very direct washback effect upon teaching curricula, it can affect the curriculum and learning (Alderson & Wall, 1993). When a test reflects the aims and objectives of the syllabus of the course, it is likely to have beneficial washback. On the other hand, when the test was at variance with the aims and the syllabus, it is likely to have harmful washback.

A curriculum is a fundamental part of EFL classes. It provides a focus on the class, and sets goals for the student. A curriculum also gives the students a guide and idea to what he/she will learn, and how he/she has progressed when the course is over. The findings from the other instrument show that the test leads to the narrowing of contents in the curriculum. It is common to claim the existence of washback, and to declare that tests can be powerful determiners, both positively and negatively, of what happens in classrooms. The findings of the syllabus and curriculum through the statistical analyses are presented in this section to examine the test's washback on teaching and learning EFL.

To avoid confusion, if a question or statement has negative wording, it is then reverse-coded. The question is coded as 'Q' for shorter presentation. The syllabus and curriculum section of the questionnaire dealt with 7 questions which addressed a number of aspects: (a) awareness of the objectives of the syllabus and curriculum (Q1), (b) appropriateness of the syllabus and curriculum (Q2),

(c) treatment and teaching of the syllabus and curriculum contents in the class (Q3, Q4, and Q5), and (d) goals of EFL curriculum and practising and testing of language skills (Q6, and Q7). A number of statistical analyses of the data were carried out to draw results. The findings are presented by themes and step by step.

5.1.2.1 The Analysis of Descriptive Statistics

Since the questions of the questionnaire survey were organised by themes, the statements discussed here are also presented by themes. Both the questionnaires (student questionnaire and teacher questionnaire) were constructed on the same domains of EFL testing and teaching. The number of questions on each domain was equal. Therefore, the findings from both questionnaires are presented and discussed simultaneously comparing the frequency and values from statistical analyses. Now, the first theme touched upon in the surveys concerns the influences of HSC examination on the syllabus and curriculum and their assumptions about the washback effects of the EFL test. Details of the findings of the student survey are also presented in the tables, histograms, and other figures. For the sake of presentation, questions are coded as student question=SQ, and teacher question=TQ.

5.1.2.1.1 Awareness of the Objectives of the EFL Curriculum

Question 1 (Q1) asked whether the participants (teachers and students) were aware of the objectives of the syllabus and curriculum. The results showed that more than 64% students (M= 2.55, STDV =.1.47) believed (strongly disagree + disagree) that they were not aware of the objectives of the syllabus and curriculum, whereas over 59% of teachers admitted (strongly disagree + disagree) that they were also not aware of the objectives of the syllabus and curriculum. The objectives of the HSC English (2000) are: to enable the learners to communicate effectively and appropriately in real-life situations, to use English effectively across the curriculum, to develop and integrate the use of the four skills of language, etc. But the teachers and their students were not aware of the objectives of the syllabus and curriculum because they only concentrated on the test and test items. They did not teach the syllabus, rather they taught to the test. It is important to mention that the HSC EFL curriculum has a set of objectives to be attained through classroom teaching. If the

202

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

teachers themselves are not aware of the objectives, it is hardly possible for them to achieve the curriculum objectives set by the state authority.   Details of the findings of Q1 are also presented in the tables numbered 5.2 and 5.3:

The frequency options are coded as Strongly Agree= SA, Agree=A, Disagree= D, and Strongly Disagree= SD

Table 5.2: Frequency counts of awareness of the objectives of the curriculum

	Significant Frequency				Negligible Frequency				
	Agreement (SA+A)		Disagreement (SD +D)		Neutral		Missing		Total
No	Freq.	Percent (%)	Freq.	Percent (%)	Freq	pct %	Freq	pct %	
SQ1	172	34.2	321	64.2	7	1.4	-	-	500
TQ1	51	40.8	74	59.2	-	-	-		125

As shown in Table 5.3, the variances among the options were 2.176 and 2.675 for the students and teachers respectively. The skewness value for student question was .520 (positive), and kurtosis value was -1.254(negative). On the other hand, the teacher skewness value was 0.312(positive), and kurtosis value was -1.609 (negative). The analysis and discussion of skewness and kurtosis values are presented separately at the latter part of this section. The histograms (Figure 5.1 and Figure 5.2) give an overall display of the findings from frequency and descriptive point of view.

Table 5.3: Descriptive statistics on awareness of the objectives of the curriculum

Qno	Mean (M)	Median	Standard Deviation (STDV)	Variance	Skewness	Kurtosis
SQ1	2.55	2.00	1.475	2.176	.520	-1.254
TQ1	2.78	2.00	1.636	2.675	.312	-1.609

The findings of the question (Q1) are presented in the histograms below (Figure 5.1 and Figure 5.2):

Figure 5.1: Awareness of the curriculum objectives (student)
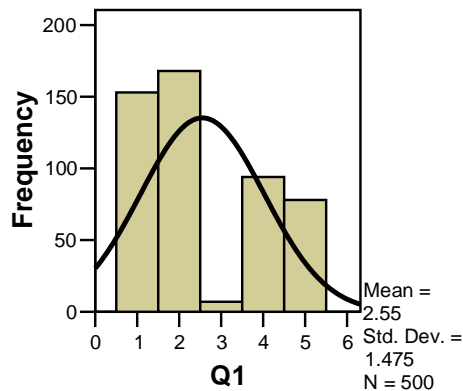
Figure 5.2: Awareness of the curriculum objectives (teacher)
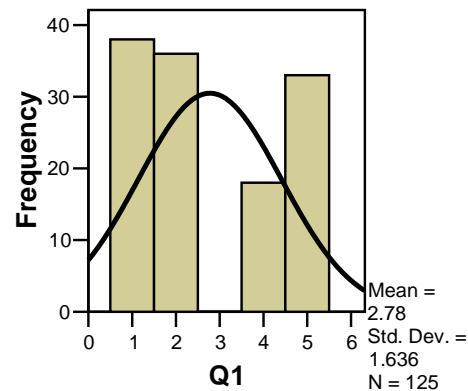
Washback has a deep relation with the syllabus and curriculum. Test contents also can have a very direct washback effect upon teaching curricula. Therefore, curriculum is a vital part of the EFL classes. Very often the test leads to the narrowing of contents in the curriculum. Tests can affect curriculum and learning (Alderson & Wall, 1993). Frontloading alignment of curriculum is commonly practiced in EFL education. A frontloaded curriculum can prevent teaching to the test, which may lead to an extremely narrow and rigid view of the actual goals and objectives of any curriculum. The findings of the study about washback onto the curriculum indicate that it operates in different ways in different situations.

The findings of the Q1 revealed that both groups of respondents were not aware of the objectives of the syllabus and curriculum. It is now strongly grounded from a number of studies that the poor knowledge of curriculum objectives is the ultimate outcome negative washback of the examination. The findings of the present study support the studies of Maniruzzaman and Hoque (2010), Maniruzzaman (2011), and Wang (2006) who find that teaching to the test and test preparation are the main concern of the teachers and their learners.

5.1.2.1.2 Appropriateness of the Syllabus and Curriculum

Q2 inquires about whether the present syllabus and curriculum enhance EFL teaching and learning. The findings of Q2 (Table 5.4 and Table 5.5) show more than 74% students (M= 3.86, STDV=1.309) and over 64% teachers (M=3.53, STDV= 1.532) suggested that the present HSC syllabus and curriculum could enhance teaching and learning:

Table 5.4: Frequency counts on appropriateness of the syllabus and curriculum

No	Significant Frequency				Negligible Frequency				Total
	Agreement (SA+A)		Disagreement (SD +D)		Neutral		Missing		
	Freq.	Percent (%)	Freq.	Percent (%)	Freq	pct %	Freq	pct %	
SQ2	371	74.2	129	24.2	8	1.6	-	-	500
TQ2	80	64	42	33.6	3	2.4	-	-	125

As a cross-referencing question, Q12 asked the participants whether the textbook, *English for Today* for classes 11-12 was well-suited for practising EFL. To this question, 61.6% teachers and 75% students were with the opinion that the textbook (which corresponded to the syllabus and curriculum) were well-suited for developing communicative competence in English.

Table 5.5: Descriptive statistics on appropriateness of the syllabus and curriculum

Qno	Mean (M)	Median	Standard Deviation (STDV)	Variance	Skewness	Kurtosis
SQ2	3.86	4.00	1.309	1.714	-.910	-.549
TQ2	3.53	4.00	1.532	2.348	-.559	-1.287

The overall objectives of the HSC English curriculum (2000) are: (a) to enable the learner to communicate effectively and appropriately in real life situations, (b) to use English effectively, (c) to develop and integrate the use of the four skills of language (listening, speaking, reading, and writing), (d) to develop an interest in and appreciation of literature, and (e) to recycle and reinforce structures already learned. The findings are also presented in the following figures (Figure 5.3 and Figure 5.4):

Figure 5.3: Appropriateness of the curriculum (student)

Mean = 3.86
Std. Dev. = 1.309
N = 500
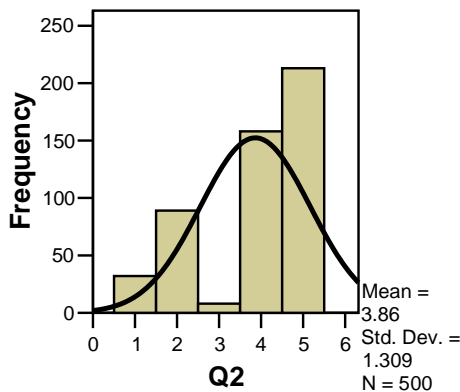
Figure 5.4: Appropriateness of the curriculum (teacher)

Mean = 3.53
Std. Dev. = 1.532
N = 125
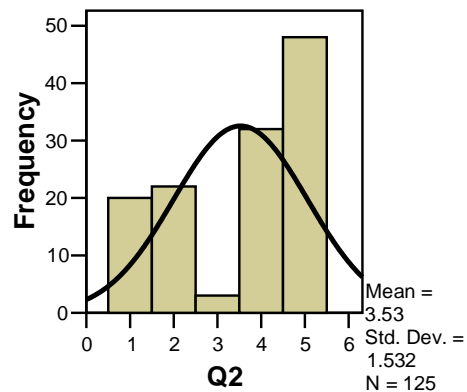
In the present study, the findings revealed that though most of the teachers and students were not aware of the objectives of the syllabus and curriculum, they believed that the syllabus and curriculum could enhance EFL learning. The view was supported by the classroom observation findings (Section 5.2).It is a well grounded fact that any curriculum cannot ensure that communicative language teaching and learning take place in the classroom. It only provides a set of criteria which, if properly implemented, would give the best possible chance for that to happen. The analysis of the syllabus and curriculum found (section 5.3) that the HSC syllabus was communicative thematically. There is a very strong question whether the set objectives of the curriculum are attainable. Because the teachers do not like to take any risk of teaching the items which are not tested, they consider it simply waste of time.

5.1.2.1.3 Teaching of the Syllabus and Curriculum

A group of questions (Q3, Q4, and Q5) asked the respondents to assess their own pedagogical knowledge and their treatment of syllabus contents in the classroom (e.g., whether they knew how to go about things in the course of their instruction and whether they were clear on the principles underpinning CLT). Q3 asked whether the teacher taught every section in the textbook whether those were tested or not. In reply to Q3, 60% teachers (M=2.52, STDV=1.501, Variance=2.252) pointed out that they did not teach every section of the syllabus. About 71% students (strongly disagree +disagree) students (M=2.37, STDV=1.376, Variance=1.894) confirmed that their teachers did not teach all the sections of the syllabus. The following tables (Table 5.6 and Table 5.7) project the findings:

Table 5.6: Frequency counts on treatment of the syllabus and curriculum

	Agreement (SA+A)		Disagreement (SD +D)		Neutral		Missing		Total Sample
	Freq.	Percent (%)	Freq	Percent (%)	Freq	pct %	Freq	pct %	
SQ3	132	26.4	352	70.8	13	2.6	3	0.6	500
TQ3	48	38.4	75	60	2	1.6	-	-	125
SQ4	360	72	129	25.8	10	2	1	0.2	500
TQ4	107	85.6	18	14.4	-	-	-	-	125
SQ5	349	69.8	146	29.2	5	1.0	-	-	500
TQ5	89	71.2	33	26.4	1	0.8	2	1.6	125

As a cross referencing question for Q3, Q11 asked whether they skipped certain sections of the textbook which were less likely to be tested in the examination. The findings of the question (Q11) directly supported the result. To the question Q11, over 70% teachers admitted that they skipped some of the sections of the syllabus, whereas nearly 75% students suggested that teachers skipped certain topics because they were less likely to be tested in the examination. The findings of Q3 are also presented in the figures below:

Figure 5.5: Teaching every section of the syllabus (student)
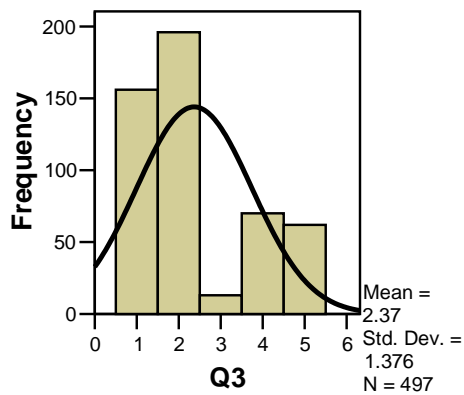
Figure 5.6: Teaching every section of the syllabus (teacher)
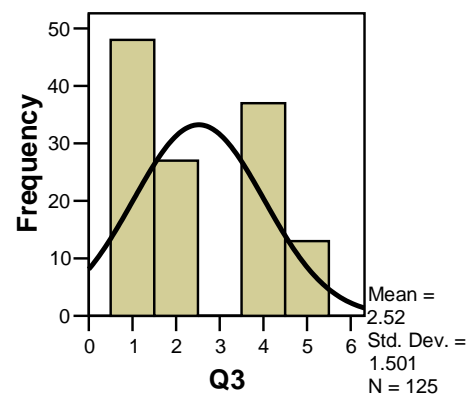
Table 5.7: Descriptive statistics on treatment of the syllabus and curriculum

Qno	Mean (M)	Median	Standard Deviation (STDV)	Variance	Skewness	Kurtosis
SQ3	2.37	2.00	1.376	1.894	.354	-.731
TQ3	2.52	2.00	1.501	2.252	.354	-1.525
SQ4	3.75	4.00	1.349	1.820	-.834	-.679
TQ4	4.17	5.00	1.183	1.399	-1.580	1.501
SQ5	3.69	4.00	1.400	1.961	-.728	-.944
TQ5	3.78	4.00	1.452	2.107	-.884	-.749

Q4 asked the respondents (teachers and the students) whether they cared about the syllabus and curriculum while preparing for the examination. The results revealed that a huge number of teachers (86%) (M=4.17, STDV=1.183, Variance= 1.399) and students (72%) (M=3.75, STDV =1.349, variance= 1.820) did not care about the syllabus and curriculum while preparing for the examination. This findings were also supported by the classroom observation (Section 5.2) using the UCOS, COLT and a self-made checklist, where the present researcher found the teachers

and the students practising the test items (model questions, and past papers) in the class. The findings are supported by Hwang (2003) who finds that learners practise the items that are tested in the examination. The figures below (Figure 5.7 and Figure 5.8) reflect the findings of this question:

Figure 5.7: Caring about the syllabus (student)
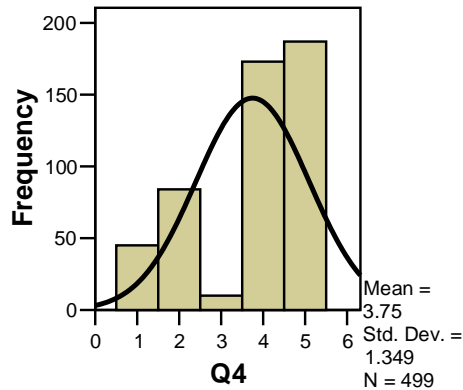
Figure 5.8: Caring about the syllabus (teacher)

Mean = 3.75
Std. Dev. = 1.349
N = 499

Mean = 4.17
Std. Dev. = 1.183
N = 125
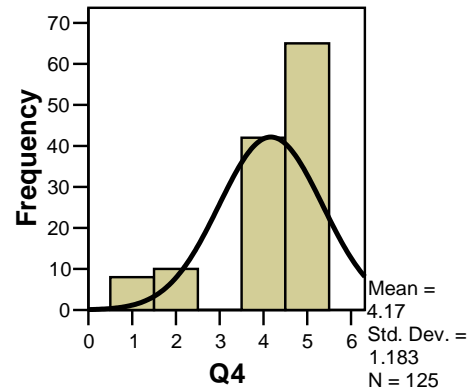
From the interview with the EFL teachers, it was found that the teachers went on their own way to prepare their students for the examination. The students also followed their teachers' instruction to prepare themselves. For Q4, Q10 can be used as a cross-referencing question which (Q 10) disclosed that majority of the respondents (75% teachers and over 66% students) believed that the students did not study the textbook materials seriously. The figures below present the frequency of responses of the students and the teachers:

Figure 5.9: Feeling pressure to cover the syllabus (students)
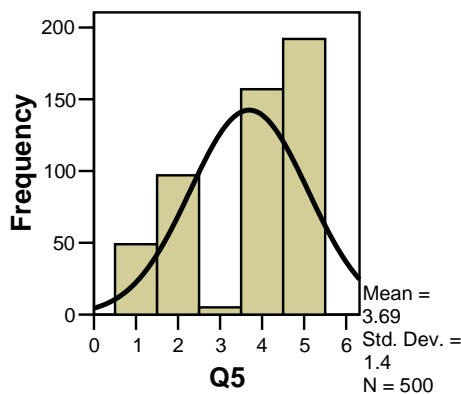
Figure 5.10: Feeling pressure to cover the syllabus (teacher)

Mean = 3.69
Std. Dev. = 1.4
N = 500

Mean = 3.78
Std. Dev. = 1.452
N = 123
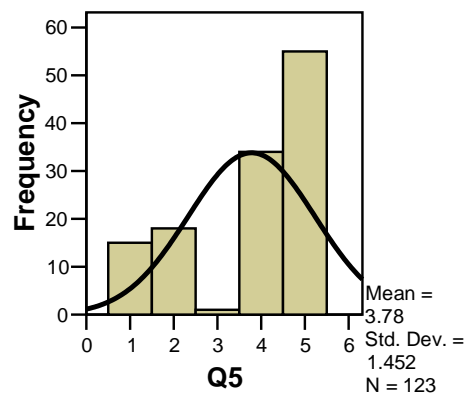
It is a well grounded fact that the students do not prefer to study the textbook materials seriously because they have alternative materials such as model questions and test papers. This classroom practice and use of commercially produced materials are the evidences of the existence of negative washback on language teaching and learning English at the higher secondary level.

Q5 asked whether the respondents felt pressure to cover the syllabus before the examination. In response to the question, nearly 70% students and more than 71% teachers pointed out that they felt pressure to complete the syllabus. It is strongly believed that a high-stakes test such as the HSC examination imposes exaggerated pressure both on teachers and students to secure good grades in the examination. This is an observable evidence of negative washback on teaching and learning English as a foreign language. It leads the teachers and students to the narrowing of the curriculum by directing teachers to focus only on those items and skills that are included in the examinations. As a consequence, such tests are said to dominate and distort the whole curriculum (Shepard, 1991).

A test is considered to have beneficial washback when it does not dominate teaching and learning activities by narrowing the curriculum. When a test reflects the aims of the syllabus of the course, it is likely to have beneficial washback, but when the test is at variance with the aims and the syllabus, it is likely to have harmful washback.

5.1.2.1.4 Goals of the EFL Curriculum and HSC Examination

Q6 asked if the HSC examination reflected the goals of HSC curriculum, that was, communicative competence. In reply to the question, over 69% students (M=2.43, STDV=1.378, Variance=1.9) and more than 59% teachers (M=2.76, STDEV=1.668, Variance=2.78) suggested that the HSC examination in English did not correspond to the objectives of the HSC English curriculum. In this study, their opinion on this question was proved by many ways: analysis of the HSC English test, analysis of the HSC answer scripts, in-depth interview with the EFL teachers and examiners, and above all the classroom observation. The tables (Table 5.8 and Table 5.9) display the detailed findings of Q6 and Q7:

209

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

Table 5.8: Frequency counts on practising and testing the competence

	Agreement (SA+A)		Disagreement (SD +D)		Neutral		Missing		Total Sample
	Freq.	Percent (%)	Freq.	Percent (%)	Freq	pct %	Freq	pct %	
SQ6	139	27.8	346	69.2	14	2.8	1	0.2	500
T Q6	50	40	74	59.2	1	0.8	-	-	125
SQ7	137	27.4	353	70.6	9	1.8	1	0.2	500
TQ7	28	22.4	96	76.8					125

Validity of the present HSC examination is found doubtful, because HSC examination does not measure what it is intended to measure. Validity relates to the extent to which meaningful inferences can be drawn from test scores (Bachman, 1990). In contrast, reliability concerns the consistency of measurement. Of the validity considerations for a language test, construct validity is viewed as pivotal.

Table 5.9: Descriptive statistics on practising and testing English

Qno	Mean (M)	Median	Standard Deviation (STDV)	Variance	Skewness	Kurtosis
SQ6	2.43	2.00	1.378	1.900	.740	-.825
TQ6	2.76	2.00	1.668	2.781	.325	-1.62
SQ7	2.37	2.00	1.365	1.864	.772	-.773
TQ7	2.12	2.00	1.435	2.058	1.111	-.261

It is often used to refer to the extent to which one can interpret a given test score as an indicator of a test takers' language ability. The term can be interpreted to mean that if a test has good construct validity, it is a good indicator of test takers' language ability and vice-versa. The histograms below (Figure 5.11 and Figure 5.25) demonstrate the comparison between teachers' and students' responses:

Figure 5.11: HSC examination and curriculum objectives (student)
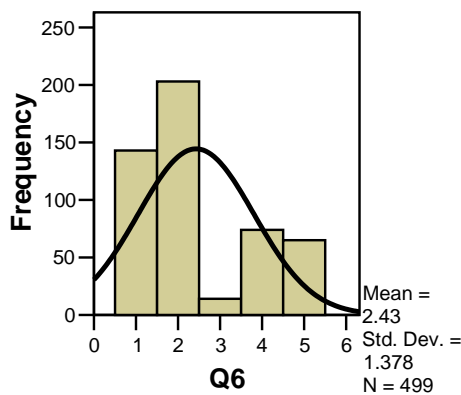
Figure 5.12: HSC examination and curriculum objectives (teacher)

Mean = 2.43
Std. Dev. = 1.378
N = 499

Mean = 2.76
Std. Dev. = 1.668
N = 125

210
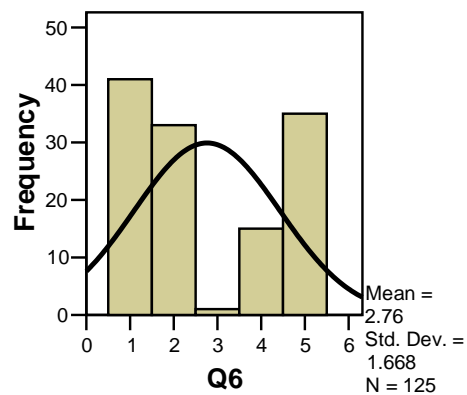
The main objective of the HSC syllabus and curriculum is to attain communicative competence, whereas the HSC EFL examination assesses mainly grammar, vocabulary, reading comprehension, and the writing skills to some extent. The last question (Q7) about the syllabus and curriculum asked if the respondents gave little attention to the examination preparation classes. In replying to the question, nearly 77% teachers (M= 2.12, STDV=1.435, Variance= 2.058) almost 70% students (M= 2.37, STDV=1.365, Variance= 1.864) disagreed with the statement meaning that they usually gave serious attention to the test items. The histograms below display the findings of the question.

Figure 5.13: Concentration on the exam preparation classes (student)
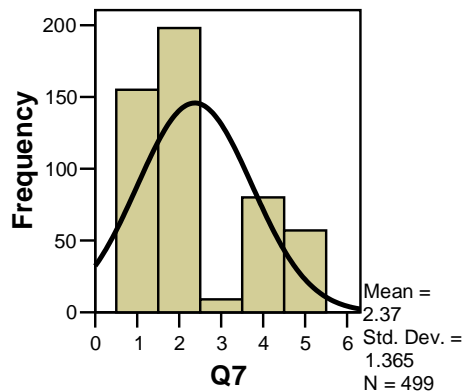
Figure 5.14: Concentration on the exam preparation classes (teacher)
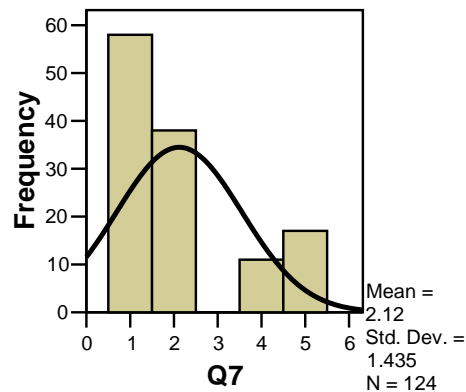
The syllabus and curriculum advocate for the communicative language teaching (CLT), but the HSC examination in English hinders the application of CLT. The study probed into the views on the impact of the EFL examination on learning (e.g., whether the test could motivate students, helped students understand their own learning needs, etc.). The results reflected that the majority of teachers showed negative impression of the impact of the HSC examination on teaching and learning EFL. Thus, the study found mismatches between teaching and testing English.

Chapman and Snyder (2000) suggest that policy makers are responsible for clarifying and elaborating the link between testing and improved teaching and learning. Although Chapman and Snyder (2000) do not articulate the role of beliefs of stakeholders, it can be argued that one of the embedded assumptions is belief change as Fullan (2001) suggests that it plays an important role in promoting desired test impact.

5.1.2.2 Skewness and Kurtosis

It was found that the findings of the student questions Q1, Q3, Q6, and Q7 (Table 5.10) had positive skewness (0.520, 0.798, 0.740, and 0.772). On the other hand, the skewness values of the questions: Q2, Q4, Q5 were negative (910, -.834, and -.727) (Table 5.10). The figure 5.15 shows how the histogram skewed positively. On the other hand, the histogram (Figure 5.17) showed how the data skewed negatively:

Table 5.10: Skewness and kurtosis value distribution (student data)

Students		SQ1	SQ2	SQ3	SQ4	SQ5	SQ6	SQ7
N	Valid	500	500	497	499	500	499	499
	Missing	0	0	3	1	0	1	1
Skewness		.520	-.910	.798	-.834	-.728	.740	.772
Std. Error of Skewness		.109	.109	.110	.109	.109	.109	.109
Kurtosis		-1.254	-.549	-.731	-.679	-.944	-.825	-.773
Std. Error of Kurtosis		.218	.218	.219	.218	.218	.218	.218

Similarly, it was observed that the findings from the teachers' (Table 5.11) questions had both positive and negative skewness. It was found that the Q7 had very highly skewed data. In addition, the findings from teacher questions: Q1, Q3, Q6, and Q7 had positive skewness values (.312, .354 .325, and 1.111); therefore, the histogram (Fig. 5.16) skewed positively. On the other hand, the teachers' questions had also negative skewness value (such as in Q2, Q4, and Q5) and the histogram skewed negatively (Figure 5.18):

Table 5.11: Skewness and kurtosis value distribution (teacher data)

Teachers		TQ1	TQ2	TQ3	TQ4	TQ5	TQ6	TQ7
N	Valid	125	125	125	125	123	125	124
	Missing	0	0	0	0	2	0	1
Skewness		.312	-.559	.354	-1.580	-.884	.325	1.111
Std. Error of Skewness		.217	.217	.217	.217	.218	.217	.217
Kurtosis		-1.609	-1.287	-1.516	1.501	-.749	-1.628	-.261
Std. Error of Kurtosis		.430	.430	.430	.430	.433	.430	.431

As mentioned, if the skewness is negative then the data is negatively skewed. For example, the histograms (Figure 5.17 and Figure 5.18) are negatively skewed.

The analysis of skewness on syllabus and curriculum both for student data and teacher data are shown in details in the tables in 5.10 and 5.11. The histograms (Figure 5.15 to Figure 5.18) display the distribution of skewness and kurtosis values. It is found that they are normally distributed:

Figure 5.15: Frequency of responses skewed positively (student)
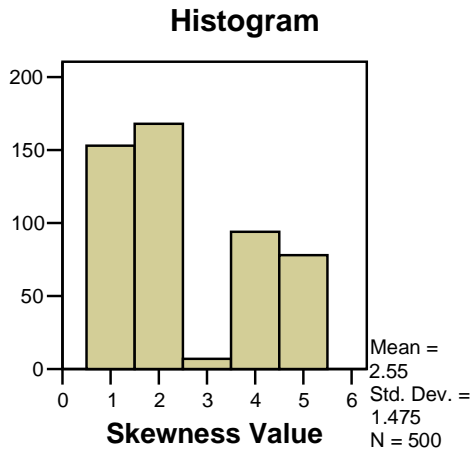
Figure 5.16: Frequency of responses skewed positively (teacher)

Histogram

Mean = 2.55
Std. Dev. = 1.475
N = 500

Histogram

Mean = 2.78
Std. Dev. = 1.636
N = 125
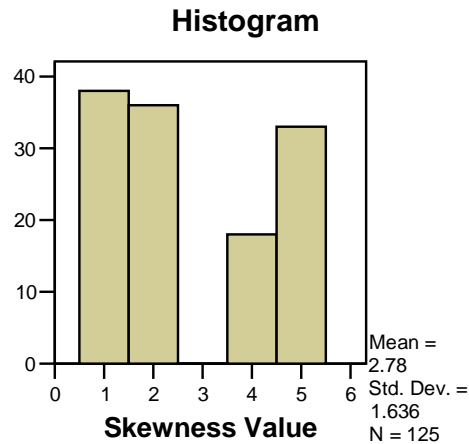
From the above discussion, it is now clear that negative skewness indicates that most of the respondents have disagreed with the statement of the question; and the positive skewness suggests that most of the respondents have agreed with statement of the question. The frequency tables (Tables 5.10 and 5.11) show the frequency of responses of agreement and disagreements of the respondents on the syllabus and curriculum. The histograms below demonstrate the skewness and kurtosis values of the questions:

Figure 5.17: Frequency of responses skewed negatively (student)
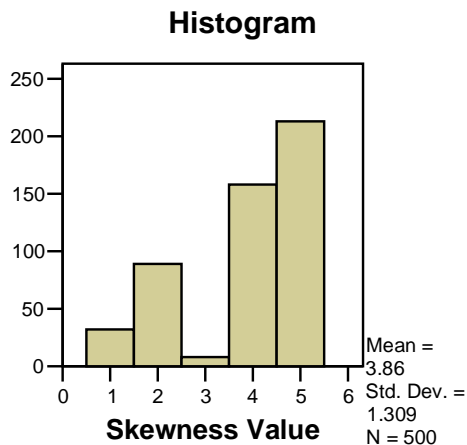
Figure 5.18: Frequency of responses skewed negatively (teacher)

Histogram

Mean = 3.86
Std. Dev. = 1.309
N = 500

Histogram

Mean = 3.53
Std. Dev. = 1.532
N = 125
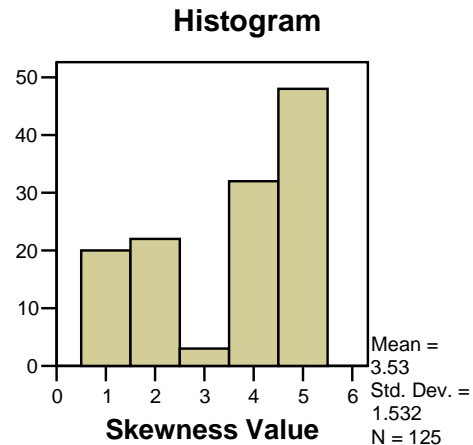
In the study, the kurtosis was calculated to observe whether the findings of questions on the syllabus and curriculum were peaked or flat relative to a normal distribution. That is, data sets with high kurtosis tended to have a distinct peak near the mean, declined rather rapidly, and had heavy tails. Data sets with low kurtosis tended to have a flat top near the mean rather than a sharp peak:

Figure 5.19: Distribution of Kurtosis results (teacher)
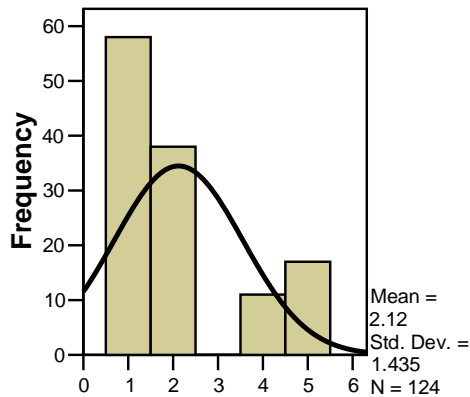
Figure 5.20: Distribution of Kurtosis results (student)

Mean = 2.12
Std. Dev. = 1.435
N = 124

Mean = 2.37
Std. Dev. = 1.365
N = 499
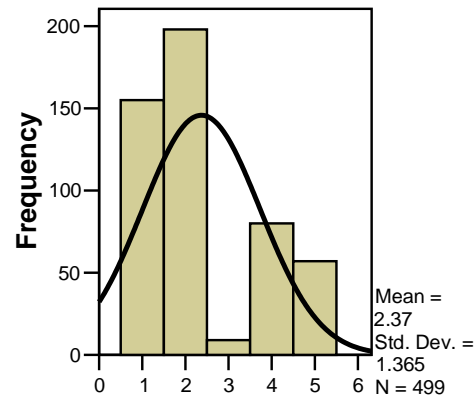
A high kurtosis distribution has a sharper peak and longer, fatter tails, while a low kurtosis distribution has a more rounded peak and shorter thinner tails. The descriptive statistics for the 7 items are presented in the above tables. The means ranged from 1.95 to 3.14 and the standard deviations ranged from 1.30 to 1.47. The medians and modes ranged from 2 to 4. The values for skewness ranged from -1.68 to 1.33, and kurtosis ranged from -.910 to 0.79. All values for skewness and kurtosis were within the accepted limits of ±3.0, indicating that the items appeared to be normally distributed:

Figure 5.21: Distribution of Kurtosis results (student)
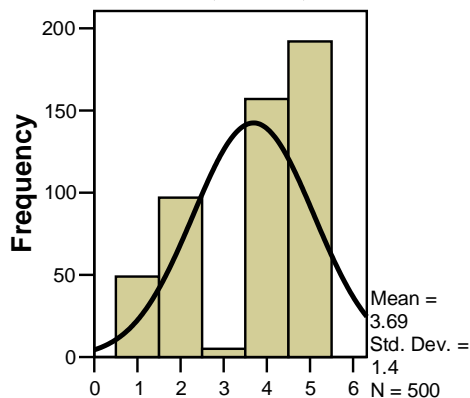
Figure 5.22: Distribution of Kurtosis results (teacher)

Mean = 3.69
Std. Dev. = 1.4
N = 500

Mean = 3.78
Std. Dev. = 1.452
N = 123
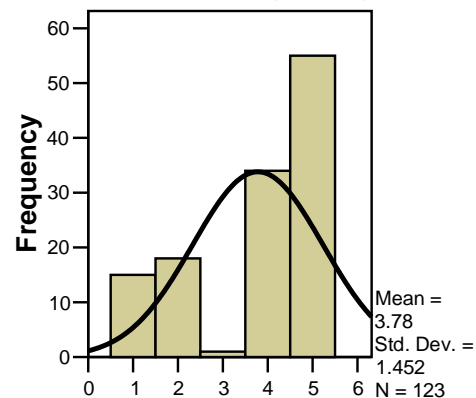
Washback has deep relation with the syllabus and curriculum. A test is considered to have beneficial washback, when preparation for it does not dominate teaching and learning activities as narrow the curriculum. When a test reflects the aims and the syllabus of the course, it is likely to have beneficial washback, but when the test is at variance with the aims and the syllabus, it was likely to have harmful washback. Test contents can have a very direct washback effect upon teaching curricula. A curriculum is a vital part of EFL classes. It provides a focus on the class and sets goals for the students throughout their study.

A curriculum also gives the student a guide and idea to what they will learn, and how they have progressed when the course is over. The test leads to the narrowing of contents in the curriculum. Tests can affect curriculum and learning (Alderson & Wall, 1993). Shohamy et al. define curriculum alignment as "the curriculum is modified according to test results" (1996, P.6). It is common to claim the existence of washback and to declare that tests can be powerful determiners, both positively and negatively, of what happens in classrooms.

5.1.2.3 The Inferential Statistical Analysis

In the previous section, the survey results from descriptive statistics concerning the principal aspects involved in the washback phenomenon–various components of the syllabus and curriculum are presented. In this section, the major research question of this study ''Does washback of the HSC public examination influence EFL teaching and learning?" is answered more extensively. Concretely, the salient findings derived from inferential statistics of the questionnaire data are now presented. Levene's Test (1960) for Equality of Variances and Independent Sample Test (T-Test) were performed for some advanced level of analysis of findings. The researcher performed the internal consistency reliability analyses to examine the homogeneity of the items.

5.1.2.3.1 Internal Reliabilities

The researcher computed internal consistency reliability estimates (i.e., coefficient alpha) of the syllabus and curriculum variables. The tables below (Table

5.12 and Table 5.13) show the reliability estimates for internal consistency of the 7 items of the questionnaire concerning the syllabus and curriculum:

Table 5.12: Reliability estimate table- (Student items)

Student items	No. Items Used	Question Number	Reliability Estimates (alpha)
Syllabus and curriculum	4 items	1, 2, 3, 4	.79
	2 items	5,6	.69
	4 items	1, 5, 6, 7	.71

For the student- items, the reliabilities of the items ranged from a low 0.69 to a relatively high 0.79 for the students' curriculum knowledge and practice.

Table 5.13: Reliability estimate table- (Teacher items)

Teacher item	No. Items Used	Question Number	Reliability Estimates Cronbach's (alpha)
Syllabus and curriculum	3 items	2, 3, 7	.78
	3 items	5,6, 4	.65
	3 items	1, 5, 6,	.73

For the teacher-items, the reliabilities ranged from .65 to a relatively high .78 In general, the reliability estimated for all the scales were relatively high or moderate. The items produced a reliability estimate of 0.79 for students and .78 for teachers, above the desirable threshold of 0.70 (Garson, 2007).

Table 5.14: Correlation coefficient between teachers and students means

Pearson's product –Moment Correlation				
Correlation coefficient	Standard Error of Coefficient	Degree of Freedom	Two tailed probability	r-squired
0.927	0.053	5	.0026*	.0086

* = significant at p < 0.05

- Pearson (r) =0.92- among the means in two group of respondents

- r$\sqrt{}$=.0086. The hypotheses for this test are: H_0: rho = 0 H_a: rho <> 0

5.1.2.3.2 Levene's Test and T-Test Analysis

The present researcher conducted Levene's test and independent sample test (T-Test) to determine whether there was a significant difference between two sets of scores. The significance level of mean differences was examined using independent sample T-Tests. The Levene's test for Equality of Variances was adopted to check the equal distribution in each subgroup. The Independent Samples Test compares the mean scores of two groups on a given variable. The following two hypotheses were dealt with to judge the significance of the difference of two independent sample groups (students and teachers):

a. Null Hypothesis- The means of the two groups are not significantly different.
b. Alternate Hypothesis- The means of the two groups are significantly different.

The two groups were independent of one another. The study compared the mean scores of HSC students with the mean scores of EFL teachers on the various issues of syllabus and curriculum to examine whether effect of the HSC examination in English affected the use and teaching the syllabus and curriculum. The Independent Samples T-test determined whether the means of the two groups were significantly different or not. Means and standard deviations of two groups are presented in the table below:

Table 5.15: Group statistics of means

Group Statistics

	Resp_type	N	Mean	Std. Deviation	Std. Error Mean
Q1	Student	500	2.55	1.475	.066
	Teacher	125	2.78	1.636	.146
Q2	Student	500	3.86	1.309	.059
	Teacher	125	3.53	1.532	.137
Q3	Student	497	2.37	1.376	.062
	Teacher	125	3.03	1.586	.142
Q4	Student	499	3.75	1.349	.060
	Teacher	125	4.17	1.183	.106
Q5	Student	500	3.69	1.400	.063
	Teacher	123	3.78	1.452	.131
Q6	Student	499	2.43	1.378	.062
	Teacher	125	2.76	1.668	.149
Q7	Student	499	2.37	1.365	.061
	Teacher	124	2.12	1.435	.129

For the first question (Q1), students' mean was 2.55 while teachers' mean was 2.78. The teachers' mean (M=2.78) was greater than the students' mean (M=2.55). It was found that the teachers' score is higher than the students' score, but the difference was only (2.78-2.55) 0.23. It proved that both groups of respondents gave almost similar responses to the awareness of the objectives of syllabus and curriculum.

Next, the Levene's Test for Equality of Variances is presented. If the Levene's Test is significant (the value under "Sig." is .05 or less than .05), the two variances are significantly different. If it is not significant (the value is greater than .05), the two variances are not significantly different; that is, the two variances are approximately equal. If the Levene's test is not significant, the second assumption should be met. In Q1, It was found that the significance was .001, which was smaller than .05. It was assumed that the variances were different. If the Levene's test is not significant, the second assumption (Equal variances not assumed) is met. The tables below (Table 5.16, Table 5.17, and Table 5.18) present the results of Levene's test and T-test of different items of syllabus and curriculum:

Table 5.16: Levene's test of equity of variances- significant deference

		Levene's Test for Equality of Variances	
		F	Sig.
Q1	Equal variances assumed	10.419	.001
	Equal variances not assumed		

* = significant at p < 0.05

Table 5.17: T-Tests for equity of means for insignificant difference

		Levene's Test for Equality of Variances		t-test for Equality of Means					95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Q1	Equal variances assumed	10.419	.001*	-1.485	623	.138	-.224	.151	-.520	.072
	Equal variances not assumed			-1.396	177.722	.165	-.224	.160	-.541	.093

* = significant at p < 0.05

Table 5.18: Findings from independent sample test

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Q1	Equal variances assumed	10.419	.001	-1.485	623	.138	-.224	.151	-.520	.072
	Equal variances not assumed			-1.396	177.722	.165	-.224	.160	-.541	.093
Q2	Equal variances assumed	17.559	.000	2.462	623	.014*	.334	.136	.068	.600
	Equal variances not assumed			2.241	171.977	.026*	.334	.149	.040	.628
Q3	Equal variances assumed	21.879	.000	-4.669	620	.000*	-.664	.142	-.943	-.385
	Equal variances not assumed			-4.291	173.871	.000*	-.664	.155	-.969	-.358
Q4	Equal variances assumed	10.467	.001	-3.191	622	.001*	-.421	.132	-.679	-.162
	Equal variances not assumed			-3.452	212.356	.001*	-.421	.122	-.661	-.180
Q5	Equal variances assumed	.036	.850	-.623	621	.533	-.088	.142	-.367	.190
	Equal variances not assumed			-.610	181.926	.543	-.088	.145	-.375	.198
Q6	Equal variances assumed	30.400	.000	-2.298	622	.022*	-.331	.144	-.614	-.048
	Equal variances not assumed			-2.052	168.852	.042*	-.331	.161	-.650	-.013
Q7	Equal variances assumed	.088	.767	1.805	621	.072	.250	.138	-.022	.522
	Equal variances not assumed			1.752	182.314	.082	.250	.143	-.032	.531

F= P<.001   *= significant at p < 0.05

For Q1, F value was 10.419 (P<.001), equal variance assumed. When the equal variance is not assumed, the Levene's test for Equality of Error Variances still offers the valid t-test. In the bottom line, the T value was -1.396 and sig. (2-tailed) was .165 which was bigger than .05 (P>.05) that is, the difference was statistically insignificant (not Significant). The findings indicated that both groups of

219

respondents gave almost same opinion that they were not aware of the objectives of the syllabus and curriculum.

For Q2, the bottom line (equal variances not assumed) was again taken. It was found value of t=-2.241, df =171.977 and the significance (2-tailed) (.026) was smaller than .05. It indicated that the difference between students' response and teachers' response was significant. The students' score was bigger than the average score of Q2, that was, students' score was greater than teachers' score –t (171.977) =2.241, P <05.

For Q3, in the upper row, the significance of F was less than .05 in the upper raw, so the t value of lower row was considered for interpretation. It was found that the significance of t in the lower row was less than .05. Here, the difference between teachers and students are negative. The t-test showed that teachers' score was bigger than students' score. In the case of Q4, the score of equal variances not assumed was considered for interpretation. The different values were: t value = -4.291, df=212.356, sig (2-tailed) = .001. It indicated that the difference between teachers' scores and students' scores was statistically significant. Therefore, the null hypothesis (the means of the two groups are not significantly different) was rejected.

For Q5, the top line (equal variances assumed) is used because the F significance was .850 which was greater than .05. The- t-value was = -.623, df=621, and the p-value = .533 (2-tailed) which was quite bigger than (P.>.05) .05. Therefore, the means difference was insignificant indicating that both groups of respondents had almost same opinion that they felt pressure to cover the syllabus before the final examination. Therefore, the alternate hypothesis was rejected.

In Q6, it was found that there was a significant difference between teachers' responses and students' responses because sig (2-tailed) was found .042 which was smaller than .05. In Q7, the top line was used for interpretation. Here, values were t=1.805, df=621, mean difference =.250, and significance of P (2-tailed) =. 072. The significance (.072) was greater than threshold value .05; therefore, t-test indicated that the difference between two means was statistically insignificant which meant that the respondents gave more attention on practising test items.

5.1.3 Textbook Materials

This section presents the findings of the perception and attitudes of the respondents towards the materials, and their real practices in the class. The assumptions about the washback effect of the HSC examination in English are also drawn from their responses. The theme of this section focused on the 11 questions (Q8- Q17) which were asked to both teachers and students. Bailey (1999 p. 30) refers to 'textbook washback' as a possible result of test use. She points out that test preparation materials are indirect evidence of washback". The appropriateness of a textbook and therefore any consideration of the possible existence of washback must be considered within the specific context in which it is being used as it might be assumed that EFL textbook content and layout vary to some extent.

Though Lam (1994) notes some innovative use of materials generated by the introduction of the revised exam (e.g. the use of teacher-produced authentic materials), he speaks of teachers as 'textbook slaves' and 'exam slaves' with large numbers of the former relying heavily on the textbook in exam classes, and of the latter relying even more heavily on past papers. He reports that teachers do this as 'they believe the best way to prepare students for examination is by doing past papers' (ibid., p. 91).

Williams (1983) points out that the importance of considering the context within which a textbook is used, writing that the textbook is a tool, and the teacher must know not only how to use it, but how useful it can be. Andrews (1994) points out that examination-specific materials end up limiting the focus of teachers and learners, and resulting in what is referred to as "narrowing of the curriculum". This term is also used by Shohamy (1992, p.514) who states that "… negative washback to programs can result in the narrowing of the curriculum in ways inconsistent with real learning and the real needs of … students". The opinion that there is the potential for texts to narrow the curriculum and encourage negative washback is also reported by Cheng (1997), Shohamy et al. (1996) and Alderson and Hamp-Lyons (1996). The findings from the descriptive and the inferential analyses are presented in the preceding pages according to themes.

5.13.1 The Descriptive Statistics

Descriptive statistics used in the study were discussed in the previous section. This section presents the findings of the survey results of the textbook materials from the quantitative data analysis. A number of statistical methods are used for data analyses. The findings are interpreted from different statistical point of view. For clear view of understanding, sufficient data tables, histograms, charts are used. Throughout the questionnaire, Five-grade Likert Scale (Likert, 1932) was used to obtain data from the respondents.

Like the previous section, 'strongly agree' and 'agree' were collapsed into 'agreement', and on the other hand 'strongly disagree' and 'disagree' were collapsed into 'disagreement' as a single category for easier discussion of the results. The histograms included mean score standard deviation, and the sample size to give a complete view of findings of each question. The skewness and kurtosis values were also deeply reflected in the histograms

The findings from the statistical analyses were grouped to have a clear look on the results. Then, the scores for individual statements belonging to each cluster were summed up. In the previous section, the survey results concerning the principal aspects involved in the washback phenomenon–various components of the survey findings on the syllabus and curriculum, and teacher practice were presented. In this part, the major research topic of this study–"how the 'materials' are manifested in the washback effect in Bangladesh context"–is discussed more broadly. The decimal numbers calculated were rounded off to the nearest whole number when the researcher reported the percentages.

5.1.3.2 Major Aspects of *English for Today* for Classes 11-12

The section includes 11 questions. Both groups of respondents were asked the same questions. Since the questions of the survey were organised by themes, the statement discussed here are also presented by themes. The internal reliability (Cronbach's alpha) of student items was 0.76 whereas the teacher items' reliability was 0.74. In the study, the researcher performed some correlation analyses on the respondents' beliefs, practice, and knowledge of textbook material (EFT). However, all of them are not presented in this thesis.

Here, five major themes related to textbook material are reported: (a) teachers' communicating the lesson's objectives with students (Q8), (b) contents and exercises of the textbook material and the washback effects of the HSC examination on their teaching and learning EFL (Q9 and Q10), (c) their narrowing down the contents of the syllabus and textbook contents, "how to teach and the ways that they teach" (Q11 and Q15), (c) their knowledge on the appropriateness of the textbook for the development of the Communicative competence, (Q12 and Q14), (d) their reliance on the test-related materials (Q13 and Q16), and (e) their use of modern equipment in the EFL class (Q17). All the themes pertained to the research questions that were posed in this study.

5.1.3.2.1 Communicating the Lesson Objectives

The first theme touched upon in the survey concerned the EFL teachers' communicating the lesson's objective, beliefs, and their assumptions about the washback effects of the HSC examination on teaching and learning. Question 8 (Q8) asked whether the teacher told the objectives of the lesson while teaching. As the tables (Table 5.19 and Table 5.20) displayed, nearly 72% students (M= 2.22, STDV=1.395, Variance=1.95) opined that their teacher did not communicate the lesson objectives, while almost 74% (M=2.25, STDV=1.441, Variance= 2.075) teachers admitted of not focusing the lesson objectives of *English for Today* (for 11 and 12) to their learners.

Table 5.19: Frequency counts on communicating the lesson's objectives

	Significant Frequency				Negligible Frequency				
	Agreement (SA+A)		Disagreement (SD +D)		Neutral		Missing		Total
No.	Freq.	Percent (%)	Freq	Percent (%)	Freq	PCT %	Freq	%	
SQ-8	129	25.8	359	71.8	11	2.2	1	0.2	500
TQ-8	32	25.6	92	73.6	1	0.8	-	-	125

It indicated that the teachers did not care about the lesson objectives set by the authority, rather their objectives were to communicating the examination instructions, which was about the preparation for the HSC examination. It is found that the evidence of washback clicked from the very first minute of starting the lesson. Both the histograms below (Figure 5.23 and figure 5.24) skewed negatively,

and the kurtosis values were normally distributed supporting the findings of the study:

Figure 5.23: Communicating the lesson's objectives (student)
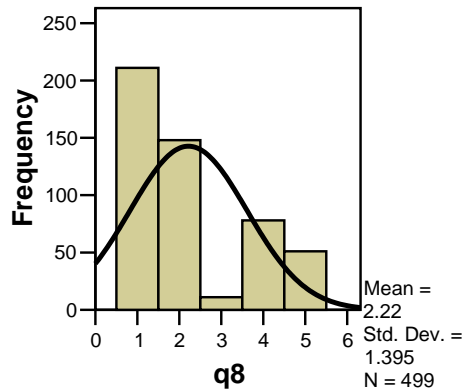
Figure 5.24: Communicating the lesson's objectives (teacher)

Mean = 2.22
Std. Dev. = 1.395
N = 499

Mean = 2.25
Std. Dev. = 1.441
N = 125
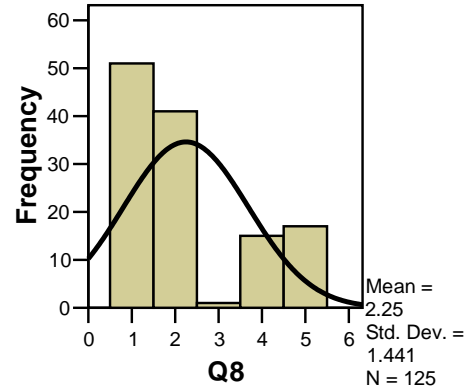
As a cross-referencing question, Q1 asked whether the participants (teachers and students) were aware of the objectives of the syllabus and curriculum. The results showed that more than 64% students and over 59% of teachers were not aware of the objectives of the syllabus. It is worth mentioning that since the textbook (*English for Today*) corresponds to the syllabus, this reference is considered to be appropriate for interpretation of the results of Q8. The table (Table 5.20) below shows the results of the descriptive statistics of Q8:

Table 5.20: Descriptive statistics on communicating the lesson objectives

Qno	Mean (M)	Median	Standard Deviation (STDV)	Variance	Skewness	Kurtosis
SQ8	2.22	4.00	1.395	1.946	-.862	-.706
TQ8	2.25	2.00	1.441	2.075	-.920	-.635

5.1.3.2.2 Contents and Exercises in *English for Today* for Classes 11-12

The issues, focused on in Q9 and Q10, were related to characteristics of the contents and exercise, and how the students treated with them. Responses to the Q9 indicated that more than 74% students (M=3.87, STDV =1.309, Variances=1.713) and 66% teachers (M=3.81. STDV = 1.479 Variance=2.189) maintained that the

textbook *English for Today* for classes 11-12 covered sufficient exercises that the syllabus and curriculum claimed (Table 5.21):

Table 5.21: Frequency counts of contents and exercises of the textbook material

| | Significant Frequency | | | | Negligible Frequency | | | | |
| | Agreement (SA+A) | | Disagreement (SD +D) | | Neutral | | Missing | | Total |
No	Freq.	Percent (%)	Freq.	Percent (%)	Freq	pct %	Freq	pct %	
SQ-9	377	75.4	114	22.8	8	1.6	1	0.2	500
TQ-9	90	72	35	28	-	-	-	-	125
SQ-10	330	66	157	31.4	11	2.2	2	0.4	500
TQ-10	93	74.4	31	24.8	-		1	0.8	125

Figure 5.25: Exercises of the textbook (student)
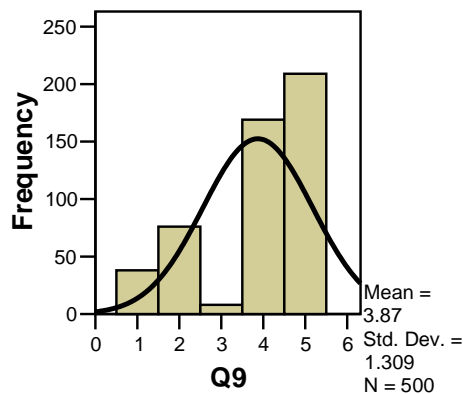
Figure 5.26: Exercises of the textbook (teacher)
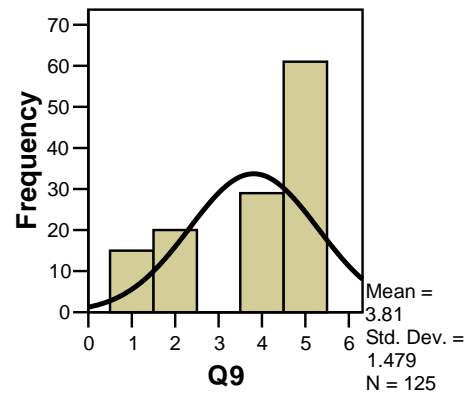
As a cross referencing question, Q2 asked whether the syllabus would enhance EFL learning. The results showed that more than 74% students and 64 % teachers believed that the present HSC syllabus and curriculum could enhance learning. Therefore, it is principled that the textbook and its contents do not influence negatively on teaching and learning, but it is the influence of examination (washback) that hinders them practising those exercises for attaining communicative competence.

The detailed findings including skewness and kurtosis are presented in table below (Table 5.22): The histograms skewed negatively because skewness values were negative. The kurtosis values were also negative with proper shape and peak (Figure 5.25 and Figure 5.26) indicating that they were normally distributed:

Table 5.22: Descriptive statistics on contents and exercises of the textbook

Qno	Mean	Median	Standard Deviation	Variance	Skewness	Kurtosis
SQ9	3.87	4.00	1.309	1.713	-.986	-.353
TQ9	3.81	4.00	1.479	2.189	-.878	-.831
SQ10	3.61	4.00	1.408	1.984	-.599	-1.127
TQ10	3.85	4.00	1.407	1.979	-.985	-.514

In response to the Q10, 66% students (M= 3.61, STDV =1.408, Variance=1.984) admitted that they did not study the textbook materials seriously while more than 74% teachers (M= 3.85, STDV =1.407, Variance=1.979) pointed out their students were reluctant in studying the textbook materials. The findings of this question can be validated through the cross-referencing by Q11 and Q13. The findings of both the questions (Q11 and Q13) supported the views of teachers and students on this (Q10) issue. The details of findings from the descriptive statistical analyses, skewness and kurtosis values are displayed in the tables (Table 5.21 & Table 5.22). The histograms (Figure 5.25 to Figure 5.28) have displayed the multilevel findings including the distribution of skewness and kurtosis values of Q9 and Q10:

Figure 5.27: Studying of the textbook materials (student)
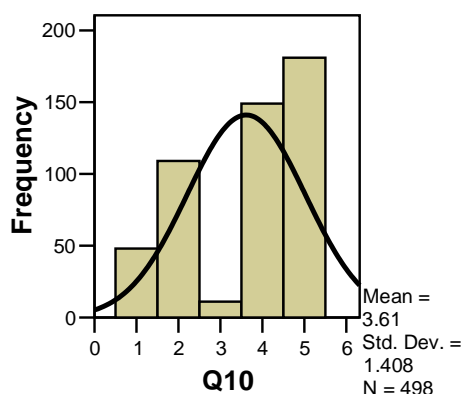
Figure 5.28: Studying of the textbook materials (teacher)
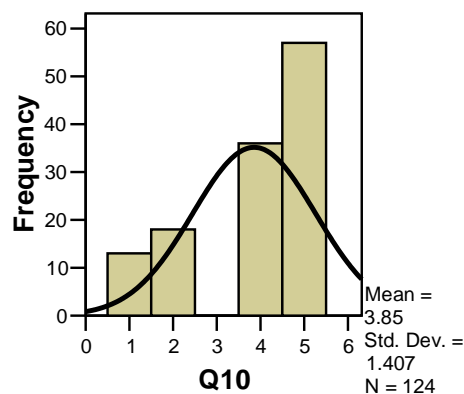
It was found that the students did not study the prescribed textbook (EFT) material because they preferred commercially produced test-related materials for the preparation of the examination. Furthermore, the students preferred to study some selected lessons and exercises likely to be tested. The findings are supported by the study of Han et al. (2004) in China. Their study finds that the teachers and the

learners are greatly dependent on commercially produced test related materials for the preparation of College English Test (CET).

Lam, (1994, p. 83) mentions that "… about 50% of the teachers appear to be "textbook slaves" in teaching the sections of the test related to listening, reading, and language systems, and practical skills for work and study". This reliance on textbooks in this context is evidence of negative washback because instead of introducing more authentic materials (the teachers) prefer to use commercial textbooks, most of which are basically modified copies of the examination paper.

5.1.3.2.3 Skipping and Narrowing the Contents of *English for Today*

The issues focused on in Q11 and Q15 were related to the respondents' views of skipping and narrowing the contents of the textbook (*English for Today* for classes-11 and 12), their concern about the test, and their role in the language classroom. Responses to Q11, nearly 75% students (M=3.85, STDV=1.29) commented that the teachers skipped certain sections of the textbook because they were unlikely to be tested in the examination. On this issue, more than 70% teachers (M=3.75, STDV=1.54) agreed with the students' claim (Table 5.24). The histograms were negatively skewed (-.912 (student) and -.831 (teacher) indicating that most of the respondents of both groups agreed with the statement of the question. The kurtosis values for students (-.511) and teachers (-.959) are negative and normally distributed (Figure 5.29 and Figure 5.30):

Figure 5.29: Skipping and narrowing the contents (student)
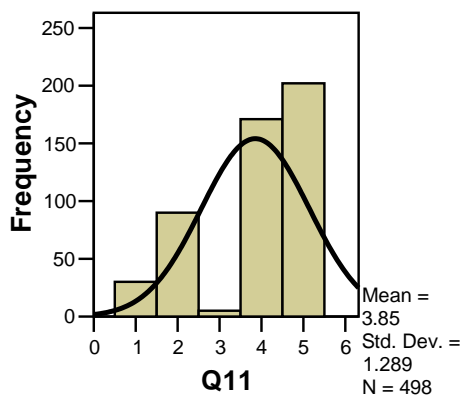
Figure 5.30: Skipping and narrowing the contents (teacher)

Mean = 3.85
Std. Dev. = 1.289
N = 498

Mean = 3.75
Std. Dev. = 1.538
N = 125
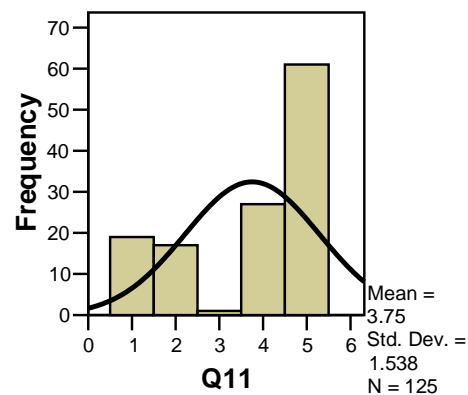
With regard to Q15, about 74% students (Table 5.23) believed that they would perform badly in the examination if they studied the whole textbook. To respond to the same question, more than 63% teachers agreed with the statement of the students on studying whole textbook (EFT). The histograms below (Figure 5.31, Figure 5.32) display the skewness and kurtosis of Q15 for both groups of respondents:

Figure 5.31: Studying of the whole textbook (student)

Mean = 3.77
Std. Dev. = 1.322
N = 495
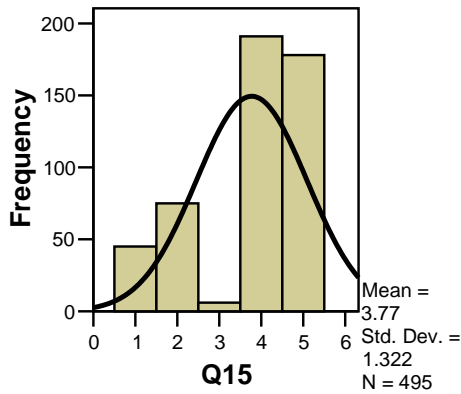
Figure 5.32: Studying of the whole textbook (teacher)

Mean = 3.77
Std. Dev. = 1.322
N = 495
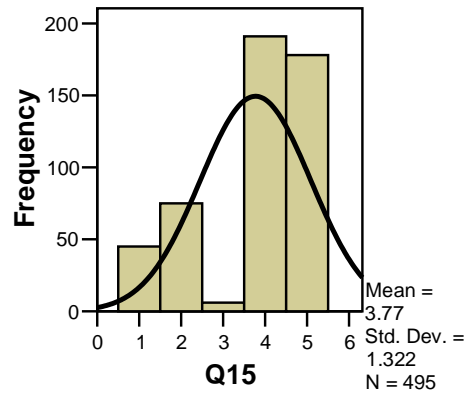
Table 5.23: Frequency counts on skipping and narrowing the contents

| Qno. | Significant Frequency | | | | Negligible Frequency | | | | Total |
| | Agreement (SA+A) | | Disagreement (SD +D) | | Neutral | | Missing | | |
	Freq.	Percent (%)	Freq.	Percent (%)	Freq	PCT %	Freq	PCT %	
SQ-11	373	74.6	120	24	5	1.0	2	0.4	500
TQ-11	88	70.4	36	28.8	1	0.8	-	-	125
SQ-15	369	73.8	120	24	6	1.2	5	1.0	500
TQ-15	79	63.2	44	35.2	2	1.6	-	-	125

The findings of both the questions were validated through a number of instruments (e.g. interview). The findings of these two questions further validated through cross referencing with the findings of Q3 and Q7 which gave almost similar results indicating that the teachers made their students practise those items which are usually tested and ignored those which were less likely to be tested in the examination. The table below (Table 5.24) presents the findings from descriptive statistics:

Table 5.24: Descriptive statistics on contents and exercises of the textbook

Qno.	Mean	Median	Standard. Deviation	Variance	Skewness	Kurtosis
SQ11	3.85	4.00	1.289	1.663	-.912	-.511
TQ11	3.75	4.00	1.538	2.365	-.831	-.959
SQ15	3.77	4.00	1.322	1.747	.110	-.466
TQ15	3.53	4.00	1.574	2.477	-.532	-1.379

From the findings described above and shown in the tables (Table 5.34 & Table 5.35), it very clear that there is enough evidence of negative washback of the HSC examination on the teaching and learning English in general and on the use of textbook (English for Today) material in particular.

5.1.3.2.4 Awareness of the Usefulness of *English for Today*

Q12 and Q14 were about the respondents' views on the characteristics and the usefulness of the textbook contents. Q12 asked whether the *English for Today* (for classes 11-12) was well-suited one to practise for developing the communicative competence. About 75% students (M=3.79, STDV =1.285, Variance=1.652) and nearly 62% teachers (M=3.39, STDV =1.550, Variance=2.402) confirmed that the textbook was well-suited one to practise for developing the communicative competence (Table 5.25 and 5.26). Referring to the Q14, almost 72% students (M=3.79, STDV =1.424) and almost 70% teachers (M=3.72, STDV =1.457) pointed out that the textbook (English for Today) contents were interesting (Table 5.25). The findings supported the claims of NCTB that the textbook, *English for Today* was written with communicative view of teaching and learning with interesting materials:

Table 5.25: Frequency counts on the characteristics of the present textbook

	Significant Frequency				Negligible frequency				
	Agreement (SA+A)		Disagreement (SD +D)		Neutral		Missing		Total
No	Freq.	Percent (%)	Freq.	Percent (%)	Freq	PCT %	Freq	PCT %	
SQ-12	374	74.8	121	24.2	4	0.8	1	0.2	500
TQ-12	77	61.6	45	36	3	2.4	-	-	125
SQ-14	358	71.6	132	26.4	7	1.4	3	0.6	500
TQ-14	87	69.6	37	29.6	1	0.8	-	-	125

Figure 5.33: Characteristics of the textbook (student)
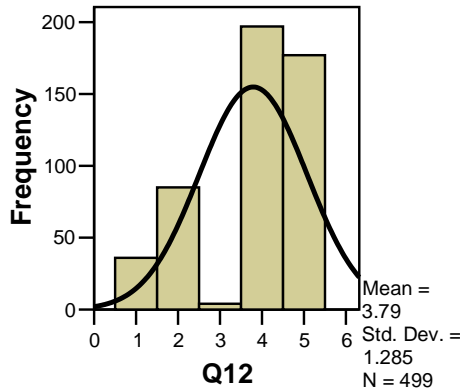
Figure 5.34: Characteristics of the textbook (teacher)
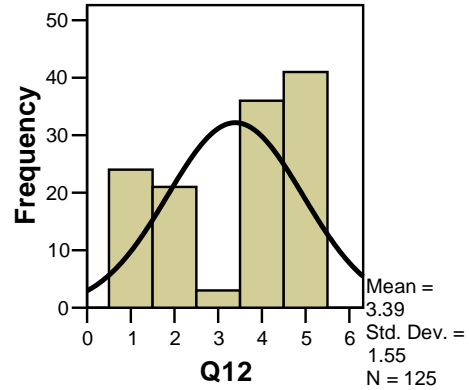
There are strong consistency and correlation in the participants' responses. For the cross-referencing of the findings, Q9 may be mentioned which asked whether the textbook covered sufficient exercises and opportunities for practising EFL. The results of this question coincided with the above two questions (Q12 and Q14). The figures (Figure 5.33 to Figure 5.36) display frequencies, means and standard deviations. The histograms are skewed negatively supporting the responses of teachers and students. The findings of these two questions are supported by Gu (2005) who finds that students give more attention when the tasks are interesting:

Figure 5.35: Quality of the textbook lessons (student)
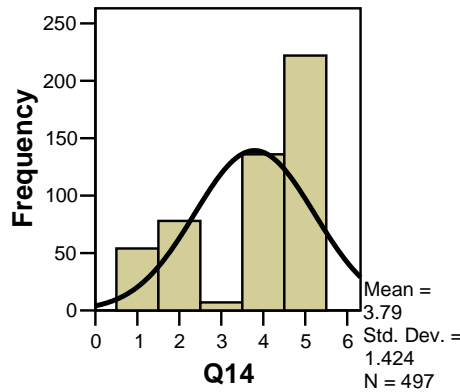
Figure 5.36: Quality of the textbook lessons (teacher)
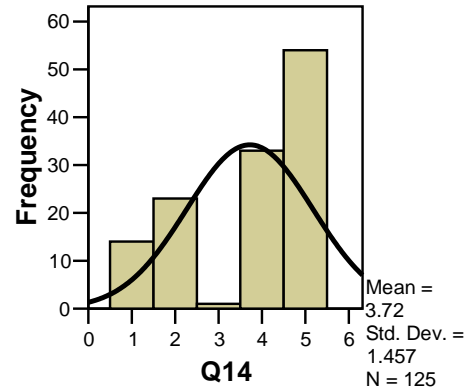
Table 5.26: Descriptive statistics on contents and exercises of the textbook

Qno	Mean (M)	Median	Standard Deviation (STDV)	Variance	Skewness	Kurtosis
SQ12	3.79	4.00	1.285	1.652	-.906	-.462
TQ12	3.39	4.00	1.550	2.402	-.455	-1.396
SQ14	3.79	4.00	1.424	2.028	-.981	.219
TQ14	3.72	4.00	1.457	2.123	-.756	-.983

The washback of the HSC examination compels the respondents to avoid practising any appropriate textbook like this, *English for Toady* (for classes 11-12). The strong evidence of negative washback is found in this section. It was found that the teachers did not teach to the contents of textbook, rather they taught to the test. Though the textbook contents were interesting, both groups of respondents preferred to practise commercially produced materials for the preparation of HSC examination. The illustrated tables (Table 5.25 and Table 5.26) project the findings in detail.

5.1.3.2.5 Types of Materials Used in the Class

Q13, Q16 and Q17 asked about the use of test- related materials and equipment used in the class. In reply to Q13, about 76% students (M=3.91, STDV =1.265, Variance=1.600) and 69% teachers (M=3.70, STDV =1.529, Variance= 2.339) disclosed that they relied on the test-related materials in the classroom for the preparation of the examination (Table 5.27 and Table 5.28). In stead of using the textbook (*English for Today*), most of the teachers were heavily dependent on the test papers, guidebooks, suggestion book, past questions, etc.

The findings are supported by the classroom observation, where the present researcher found that nearly 80% EFL teachers used commercially produced guide books, test papers, model questions, etc. The findings are further validated by the findings of the interviews with the EFL teachers. The interviewed teachers revealed that they used test related guide book, suggestion book, etc. to prepare their students. The tables and figures below demonstrate the detailed findings of the 3 questions:

Table 5.27: Frequency counts on the types of materials used

| | Significant Frequency | | | | Negligible Frequency | | | | |
| | Agreement (SA+A) | | Disagreement (SD +D) | | Neutral | | Missing | | Total |
Qno.	Freq.	Percent (%)	Freq.	Percent (%)	Freq	PCT %	Freq	PCT %	
SQ-13	378	75.6	110	22	10	2.0	2	0.4	500
TQ-13	86	68.8	38	30.4	1	0.8	-	-	125
SQ-16	31	24.8	92	73.6	2	1.6	-		125
TQ-16	89	17.8	398	79.6	11	2.2	2	0.4	500
SQ-17	112	22.4	376	75.2	8	1.6	-	-	500
TQ-17	17	13.6	106	84.8	2	1.6	-	-	125

Figure 5.37: Reliance on test-related materials (student)
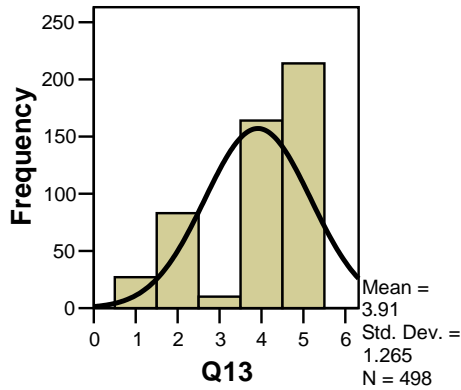
Figure 5.38: Reliance on test-related materials (teacher)
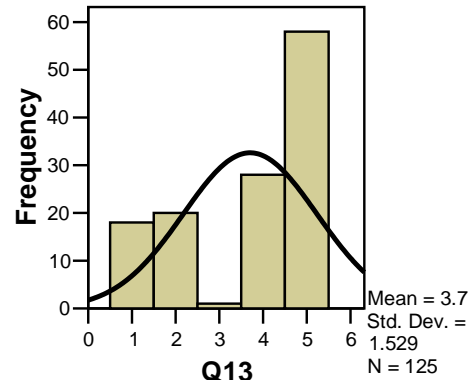
This findings are again supported by Cheng (1997, p.50) who notes the existence of workbooks specifically designed to prepare students for examination papers in the Hong Kong Certificate of Education Examination and the heavy reliance of teachers on these workbooks. On the topic of textbook evaluation, Williams (1983, p.254) highlighted the importance of considering the context within which a textbook is used.

For cross- referencing of the findings, the results of the Q3, Q7, and Q27 were checked and coordinated; and it was found that the findings were valid. The reliability of the findings was once again proved by a number of ways such as the classroom observation, and interview with the teachers. During the classroom observation the researcher found that most of the observed teachers (80%) used the direct test related materials in their class while 30% observed teachers did not bring the original textbook for the practice in the class:

Table 5.28: Descriptive statistics on the types of materials used

Qno	Mean	Median	Standard. Deviation	Variance	Skewness	Kurtosis
SQ13	3.91	4.00	1.265	1.600	-.981	-.343
TQ13	3.70	4.00	1.529	2.339	-.752	-1.071
SQ16	2.08	2.00	1.241	1.541	1.213	.385
TQ16	2.26	2.00	1.442	2.079	.935	-.595
SQ17	2.05	1.00	1.418	2.011	1.098	-.315
TQ17	1.96	2.00	1.174	1.377	1.508	1.481

Figure 5.39: Use of authentic
materials (student)

Mean =
2.08
Std. Dev. =
1.241
N = 498
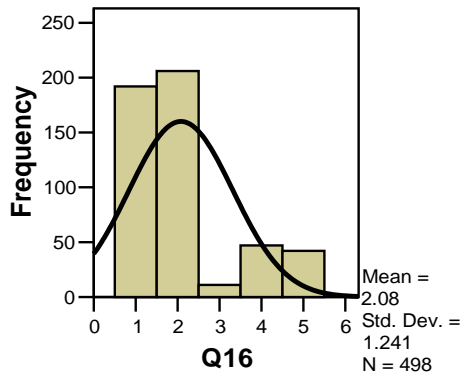
Figure 5.40: Use of authentic
materials (teacher)

Mean =
2.26
Std. Dev. =
1.442
N = 125
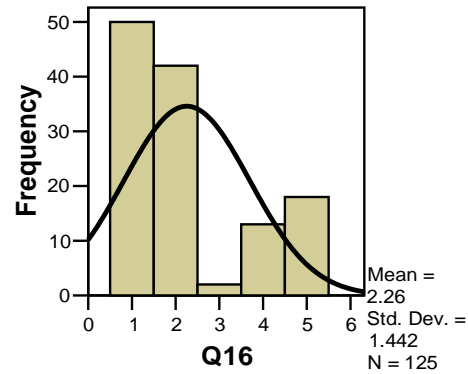
For Q16, nearly 74% students (M=2.26, STDV = 1.442, Variance= 1.241) and 80% teachers (M= 2.26, STDV = 1.442, variance= 2.079) commented that teachers did not use authentic materials in the classroom. Authenticity is very important, as Bailey (1996, p.276) referring to a test promoting positive washback states, "… a test will yield positive washback to the learner and to the programme to the extent that it utilises authentic tasks and authentic texts".

For the Q17, approximately 85% teachers admitted that they did not use any modern equipment in the class while more than 75% students supported their teacher's statement of not using the modern technology in the language classroom (Table 5.27). The histograms (Figure 5.39 to Figure 5.42) display the detailed findings along with the projection of skewness and kurtosis values. The figures display that they are properly skewed and shaped with properly distribution. Hoque (2008) reveals that the EFL teachers at the HSC level used only board, chalk, and textbook for teaching English as a foreign language.

Figure 5.41: Use of modern
equipment (student)

Mean =
2.05
Std. Dev. =
1.418
N = 496
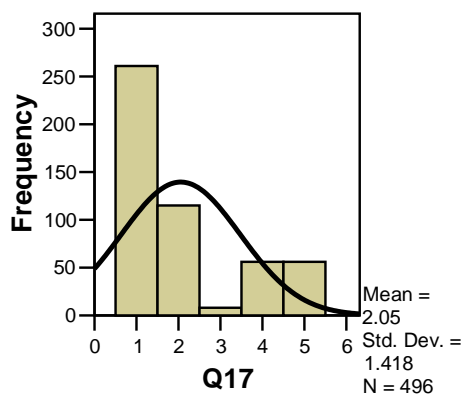
Figure 5.42: Use of modern
equipment (teacher)

Mean =
1.96
Std. Dev. =
1.174
N = 125
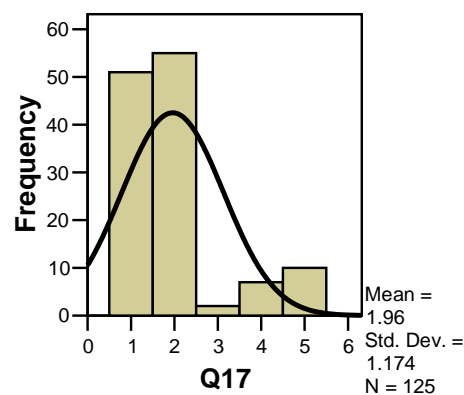
The CLT classroom requires equipment and technology, but there is no facility of using modern technology such as multimedia projector, overhead projectors and the like in language classrooms at the higher secondary level in Bangladesh. The twenty-first century is the age of modern technology, globalization and changes which have a great impact on teaching and learning. Foreign language (FL) teachers have now been leaders in the use of technology in the classroom, from short wave radio and newspapers, to film strips, to tape recorders, to records, films, video, computers, multimedia, and now internet, as a means of bringing authentic language and culture to their students.

The Twenty- first century foreign language teachers must learn to use technology effectively and meaningfully. But in Bangladesh, the language teachers teach English through the Grammar Translation Method, the language teachers heavily rely on the textbooks, wall boards and other traditional teaching aids and equipment. As a result, the teachers' effort becomes unsuccessful in teaching English language, a very vibrant and living language. Now- a- days, modern technologies are found available in some urban colleges, but due to lack of technical training and experience, the teachers cannot use them in their English language classes.

5.1.3.3 Internal Reliability

The internal reliability of questions for every section (e.g., syllabus and curriculum, textbook materials, teaching methods, etc) had been performed separately. Like the previous section, the researcher computed internal consistency reliability estimates (i.e., coefficient alpha) of textbook materials variables.

The internal reliability of the items was estimated, and it was found that the Cronbach's alpha value ranged from 0.69 to 0.81 (group reliability 0.74) for the teacher items (Table 5.28) and ranged from 0 .68 to 0.77 (group reliability 0.76) for student items (Table 5.30). The reliability estimates for the scale were relatively high. It was assumed that the reliability of the items was significant for both groups; therefore, the findings were valid and applicable, above the desirable threshold of 0.70 (Garson, 2007):

Table 5.29: Internal consistency reliability (teacher items)

	No. of Items used	Question	Reliability Estimates (alpha)
Reliability Estimates for 'Textbook Materials			
1	3 items	8,10,13	0.71
2	5 items	9, 12,14, 17	0.73
3	3 items	10,16,17,	0.69
4	3 items	13,15,16	0.81
Reliability (Cronbach's alpha)			=0.74

Table 5.30: Internal consistency reliability (student items)

	No. of Items used	Question	Reliability Estimates (alpha)
Reliability Estimates for 'Textbook Materials -Cronbach's alpha			
1	4 items	8,9,13 , 14	0.68
2	4 items	9, 10,14, 17	0.71
3	5 items	11,12, 13, 16,17,	0.77
Reliability (Cronbach's alpha)		= 76	

5.1.3.4 T-Test Analysis of Textbook Materials

The researcher performed Levene's test and independent samples test to examine whether the means of two groups (teachers and students) responded to the questions regarding textbook materials normally. The independent samples t-test compares the mean scores of two groups on a given variable. For the independent samples T-Test, it is assumed that both samples come from normally distributed samples with equal standard deviations (or variances). In the table (5.31) below, the means of two groups are presented to compare the scores on the particular textbook materials sections. The present study applied the MMR approach for data collection and data analysis; the findings from the descriptive statistics on this area are already presented.   In this section, the findings from inferential statistics are presented to show whether washback effect influenced the teaching and learning English in general and textbook materials in particular. This domain included 10 questions of different issues; the findings of those questions are analysed through Levene's Test for Equality of Variances as well. The group statistics of textbook material are presented in the table below:

Table 5.31: Group statistics of means on textbook materials

	Resp_type	N	Mean	Std. Deviation	Std. Error Mean
Q8	Student	499	2.22	1.395	.062
	Teacher	125	2.25	1.441	.129
Q9	Student	500	3.87	1.309	.059
	Teacher	125	3.81	1.479	.132
Q10	Student	498	3.61	1.408	.063
	Teacher	124	3.85	1.407	.126
Q11	Student	498	3.85	1.289	.058
	Teacher	125	3.75	1.538	.138
Q12	Student	499	3.79	1.285	.058
	Teacher	125	3.39	1.550	.139
Q13	Student	498	3.91	1.265	.057
	Teacher	125	3.70	1.529	.137
Q14	Student	497	3.79	1.424	.064
	Teacher	125	3.72	1.457	.130
Q15	Student	495	3.77	1.322	.059
	Teacher	125	3.53	1.574	.141
Q16	Student	498	2.08	1.241	.056
	Teacher	125	2.26	1.442	.129
Q17	Student	496	2.05	1.418	.064
	Teacher	125	1.96	1.174	.105

In Q8 (Table 5.32), the Levene's value (F) is = .788 (sig) which was greater than P=.05; therefore, top line parameter was used for the T-Test results interpretation. Here, t=-.210, df =622, mean differences=-.030, and P=.833 (P>0.05) which was greater than standard value of 0.05. The Levene's test revealed that the difference between two groups was not statistically significant:

Table 5.32: Levene's test of equity of variances- significant deference

		Levene's Test for Equality of Variances	
		F	Sig.
Q8	Equal variances assumed	.072	.788
	Equal variances not assumed		

*= significant at p < 0.05

It was found that students' mean (M=2.22) and teacher's mean (M=2.25) were very close, and the difference was negligible. It indicated that both groups of respondents admitted that the textbook lesson objectives were not communicated with the students. The table (Table 5.33) below shows how the deference between two groups of respondents was insignificant:

Table 5.33: Levene's test for equality of variances

		Levene's Test for Equality of Variances		T-Test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Q8	Equal variances assumed	.072	.788	-.210	622	.833	-.030	.140	-.305	.246
	Equal variances not assumed			-.206	186.545	.837	-.030	.143	-.312	.253

*= significant at p < 0.05

The T-Test results were supported by a number of ways: questionnaire survey, classroom observation, and in-depth interview. Levene's Test for Equality of Variances (Sig. is greater than .05) in Q9, the two variances were not significantly different; that was, the two variances were approximately equal. Here, the students' mean (M=3.87), teachers' mean (M=3.87) were nearly equal. The values were: t value = .461, df=623, and p=.645, sig, (2-tailed). Here, the significance (p) (2-tailed) is 0.645 which was greater than standard level (p<0.05). The findings indicated that the difference between teachers' mean and students' mean was statistically insignificant. The findings accepted the null hypothesis of homogeneity. Details of findings are presented in the table (Table 5.34) below:

Table 5.34: Results of the independent samples test

		Levene's Test for Equality of Variances		T-Test for Equality of Means					95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Q8	Equal variances assumed	.072	.788	-.210	622	.833	-.030	.140	-.305	.246
	Equal variances not assumed			-.206	186.545	.837	-.030	.143	-.312	.253
Q9	Equal variances assumed	7.594	.006	.461	623	.645	.062	.134	-.202	.326
	Equal variances not assumed			.429	175.588	.669	.062	.145	-.224	.348
Q10	Equal variances assumed	2.043	.153	-1.701	620	.089	-.240	.141	-.518	.037
	Equal variances not assumed			-1.702	189.159	.090	-.240	.141	-.519	.038
Q11	Equal variances assumed	14.495	.000*	.755	621	.451	.101	.134	-.162	.365
	Equal variances not assumed			.680	170.290	.498	.101	.149	-.193	.396
Q12	Equal variances assumed	25.694	.000*	2.961	622	.003*	.398	.134	.134	.661
	Equal variances not assumed			2.649	169.178	.009	.398	.150	.101	.694
Q13	Equal variances assumed	19.730	.000*	1.585	621	.113	.210	.132	-.050	.469
	Equal variances not assumed			1.416	168.997	.159	.210	.148	-.083	.502
Q14	Equal variances assumed	.646	.422	.508	620	.611	.073	.143	-.208	.354
	Equal variances not assumed			.501	188.025	.617	.073	.145	-.214	.359
Q15	Equal variances assumed	23.365	.000*	1.769	618	.077	.244	.138	-.027	.514
	Equal variances not assumed			1.595	170.753	.113	.244	.153	-.058	.545
Q16	Equal variances assumed	10.840	.001*	-1.383	621	.167	-.178	.128	-.430	.075
	Equal variances not assumed			-1.265	172.941	.208	-.178	.140	-.455	.100
Q17	Equal variances assumed	16.950	.000*	.687	619	.492	.094	.137	-.175	.364
	Equal variances not assumed			.769	224.434	.443	.094	.123	-.148	.336

*= significant at p < 0.05

The difference was again insignificant in the case of Q10; here, Levene's significance (sig) was .153, quite bigger than .05. So, the Equal variances assumed (top line) was used for interpretation. In this, the values included t =-1.701, df=620, means difference = -.240, and p=.089 (greater than .05). It indicated that the two variances were statistically insignificant. For Q11, Q12, and Q13, Equal variances not assumed (bottom lines) were used for t value interpretation because Levene's significances (sig) in those questions were less than .05. In Q11, the parameters were: t= .680, df= 170.290, mean difference=.101 and p=.498. It indicated that the variances were not significantly different.

In Q12, the findings were: t=2.649, df=169.178, mean difference=.398, and p=.009 (p>.05). It suggested that two variances were significantly different because the p=.009 was smaller than .05. It was also found that the mean difference (.398) was relatively big. In reply to this question (Q12), nearly 75% students suggested that the textbook, *English for Today* (for classes 11-12) was well suited for developing the communicative competence, whereas more than 61% teachers believed that the textbook was well- suited. It was found that the variances in the mean score in Q13 were statistically different. For this case, the results were t=1.416, df=168.997, and p=.159 (sig, 2-tailed). The significance of p (p >0.05) is bigger than desirable threshold   at the significance level 0.05 which was greater than .159; it was assumed that variances were significantly different. So, the null hypothesis was rejected here.

For the case of Q14, 'Equal variances assumed' (top line) parameter was used to determine the variances. The top line parameter was taken because the Levene's significance (sig) was greater than .05. Here, the findings were: F=.646, Levene's significance (F) = .422, t=.508, and p=.611(2-tailed). The t significance (2-tailed) was greater (P>.05) than .05. So, it was concluded that the responses of teachers and students were not significantly different indicating that the learners found interest in studying the present EFL textbook, *English for Today* (for classes 11-12).

For the case of Q15, Levene's significance was less than .05 (P<.05); therefore, 'Equal variances not assumed' was used. Here, the findings were: t =1.595, df =170.753, mean difference =.244, and most importantly, p=.113. The significance (.113) (2-tailed) was greater than .05. It indicated that   variances were

not significantly different (insignificant).   The findings showed that the variance in Q16 was insignificant. The findings were: t=-1.265, df=224.434, mean difference=-.178, and p=.208. The p =.208 were greater than .05 indicating that the mean differences were insignificant. Similarly, in Q17, it was found that t= .769, df=224.434, and the mean difference (.094) between teachers and students on this issue was statistically insignificant. Furthermore, the significance (2- tailed) = .443 which was greater than standard threshold level (p<0.05). It was found that both groups of respondents largely agreed that modern equipment was not used in the language class. For this case, the null hypothesis was accepted indicating that the difference of responses was insignificant.

5.1.4 The Teaching Methods and Approaches

Andrews et al. (2002) point out that the exam leads to teachers' use of explanation of techniques for engaging in certain exam tasks. Cheng (1997) suggests that teaching methods may remain unchanged even though activities change as a result of the revision of an exam (p, 52) while Alderson and Wall's (1993, p. 127) Sri Lanka study shows the exam 'had virtually no impact on the way that teachers teach'. The high-stakes EFL exam such as the HSC examination in English leads teachers to teach through simulating the exam tasks or through carrying out other activities that directly aim at developing exam skills or strategies. Watanabe's (2000, p. 45) findings for this area are once again significant. He reports that the teachers in his study 'claimed that they deliberately avoided referring to test taking techniques, since they believed that actual English skills would lead to students' passing the exam'.

Some of the studies (e.g. Hwang, 2003) indicate that the methods used to teach towards exams vary from teacher to teacher. Alderson and Hamp Lyons (1996), and by Watanabe (1996) find large differences in the way teachers teach towards the same examination or examination skills, with some adopting much more overt 'teaching to the test', 'textbook slave' approaches, while others adopted more creative and independent approaches (p, 292). They discuss various teacher-related factors that may affect why and how a teacher works towards an examination for attaining high scores.

Teacher attitude towards an examination would seem to play an important role in determining the choice of methods used to teach exam classes. There has been a perception that washback affects teaching content and teaching methods. Other findings on teaching methods relate to interaction in the classroom. Alderson and Hamp Lyons (1996) note that the examination classes spend much less time on pair work, that teachers talk more and students less, that there is less turn taking, and the turns are somewhat longer. Watanabe (2000, p. 44) notes that 'students rarely asked questions even during exam preparation lessons'. Cheng (1998) points out that while teachers talk less to the whole class as a result of the revised exam, the teacher talking to the whole class remains the dominant mode of interaction.

5.1.4.1 Descriptive Statistics

This section presents the findings of the survey results of the teaching methods and approaches the teachers apply while teaching EFL at the HSC level in Bangladesh. The findings of this section follow the similar styles as applied in the previous sections of the findings. The findings from frequency counts, mean score, standard deviation, skewness, kurtosis, etc for each question are presented step by step. The histograms are used to focus mean score, standard deviation, skewness, and kurtosis of every individual finding. One of the crucial benefits of using histograms is that they can display a number of statistical results simultaneously. The findings have also been shown in a number of tables. They are derived from inferential statistical analyses of the questionnaire data and are presented in next section.

5.1.4.2 Major Aspects of the Methods and Approaches

The section includes altogether 9 questions on different issues on the use of teaching methods in the class. The questions in this section different aspects of teaching methods used in the class are: (a) teachers' care on students' understanding (Q18), (b) teachers' language of instructions (Q19, Q22, and Q24), (c) teachers' encouragement and motivation (Q20, and Q21), (d) teaching to the test (Q23, and Q25), and (e) indication of examination results.

5.1.4.2.1 Teachers' Care of Students' Understanding

In response to Q18, over 75% students (M=3.92, STDV=1.346, Variance=1.813) and more than 66% teachers (M=3.63, STDV=1.516, Variance=2.299) confirmed that teachers took care of their students whether they (students) understood teacher's instruction. The tables below (Table 5.35 & Table 5.36) display the findings of this question:

Table 5.35: Frequency counts on teacher's care for students' understanding

| | Significant Frequency | | | | Negligible Frequency | | | | |
| | Agreement (SA+A) | | Disagreement (SD +D) | | Neutral | | Missing | | Total |
No	Freq.	Percent (%)	Freq.	Percent (%)	Freq	pct %	Freq	pct %	
SQ18	376	75.2	112	22.4	11	2.2	1	0.2	500
TQ18	83	66.4	39	31.2	3	2.4	-	-	125

Table 5.36: Descriptive statistics on teacher's care for students' knowledge

Qno	Mean (M)	Median	Standard Deviation (STDV)	Variance	Skewness	Kurtosis
SQ18	3.92	4.00	1.346	1.813	-1.033	-.338
TQ18	3.63	4.00	1.516	2.299	-.666	-1.153

The histograms below (Figure 5.43 and 5.44) show that they are negatively skewed (to the left) meaning that most of the students and teachers showed their agreement towards the statement of the question. As shown in the histograms, the scores of kurtosis of this question for both groups are also negative (Table 5.36). They are frequent modestly-sized deviations with normal peaks and flats:

Figure 5.43: Teachers' care (student)   Figure 5.44: Teachers' care (teacher)
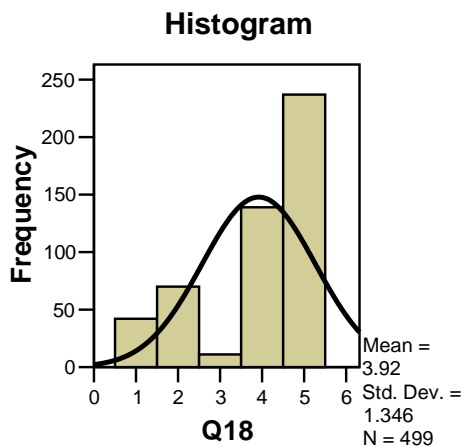
Histogram

Mean = 3.92
Std. Dev. = 1.346
N = 499

Histogram

Mean = 3.63
Std. Dev. = 1.516
N = 125
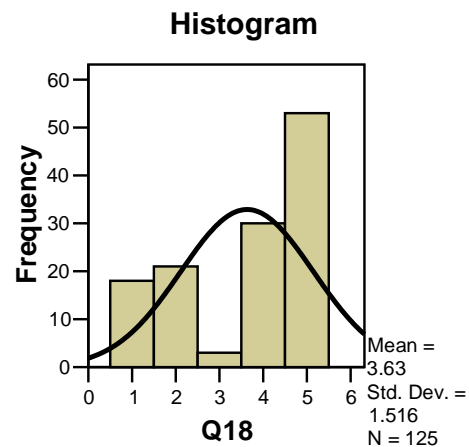
5.1.4.2.2 Teachers' Language of Instruction

Q19, Q22, and Q24 asked about the explanation of text and the medium of instructions in the class. For question Q19, nearly 50% teachers (M=3.10, STDV=1.527, Variance=2.332) and more than 66% students (M=2.52, STDV=1.455, Variance=2.117) confirmed that the teachers did not explain the text in English. Q22 was used as a cross-referencing question, where nearly 69% teachers (M=3.73, STDV=1.472) as well as over 72% students (M=3.89, STDV=1.319) indicated that teachers used Bengali along with English as the languages of instructions in the class. The findings of this section are presented in the tables below (Table 5.37 and Table 5.38):

Table 5.37: Frequency counts on teacher's care for students' understanding

| | Significant Frequency | | | | Negligible Frequency | | | | |
| | Agreement (SA+A) | | Disagreement (SD +D) | | Neutral | | Missing | | Total |
No	Freq.	Percent (%)	Freq.	Percent (%)	Freq	pct %	Freq	pct %	
SQ19	155	31	331	66.2	11	2.2	3	0.6	500
TQ19	61	48.8	62	49.6	1	.8	1	.8	125
SQ22	371	74.2	110	22	11	2.2	8	1.6	500
TQ22	86	68.8	39	31.2	-	-	-	-	125
SQ24	362	72.4	130	26	6	1.2	2	.4	500
TQ24	76	60.8	47	37.6	2	1.6	-	-	125

Figure 5.45: Explanation of text (student)
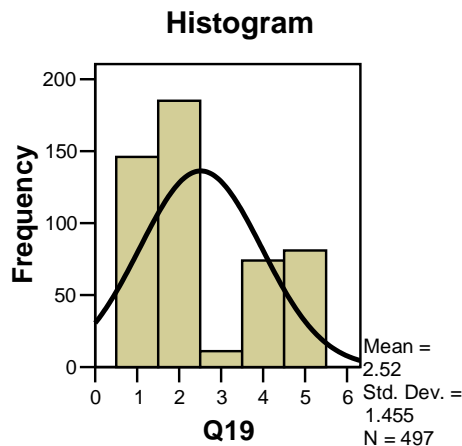
Figure 5.46: Explanation of text (teacher)

Histogram

Mean = 2.52
Std. Dev. = 1.455
N = 497

Q19

Histogram

Mean = 3.1
Std. Dev. = 1.527
N = 124

Q19
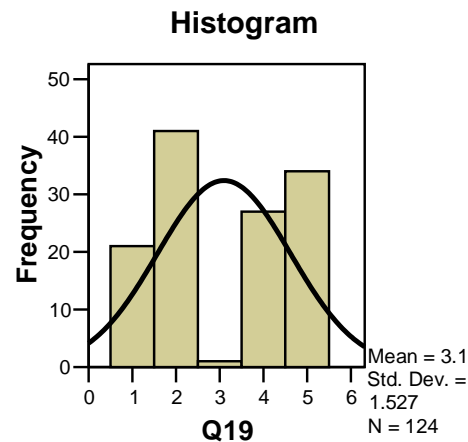
The figures above show both the histograms skewed positively (skewness, student=.626, teacher=.015) indicating that major parts of respondents disagreed with the statement of question (Q19). The kurtosis (student = -1.092, teacher = -1.616) is also significantly distributed in proper shape:

Table 5.38: Descriptive statistics on teacher's instructions of language

Qno	Mean	Median	Standard Deviation	Variance	Skewness	Kurtosis
SQ19	2.52	2.00	1.455	2.117	.626	-1.092
TQ19	3.10	2.50	1.527	2.332	.015	-1.616
SQ22	3.89	4.00	1.319	1.741	-1.006	-.331
TQ22	3.73	4.00	1.472	2.167	-.720	-1.090
SQ24	3.80	4.00	1.4	1.958	-.881	-.697
TQ24	3.14	4.00	1.50	2.231	.414	-1.436

Figure 5.47: Language of instruction (student)
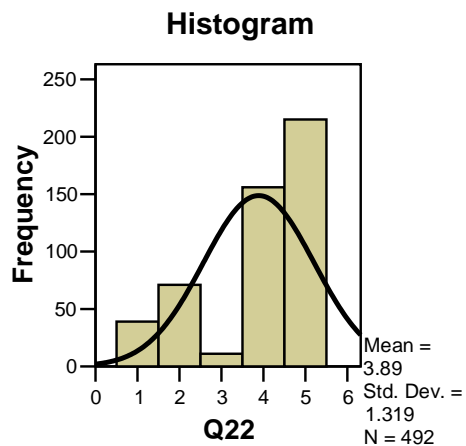
Figure 5.48: Language of instruction (teacher)
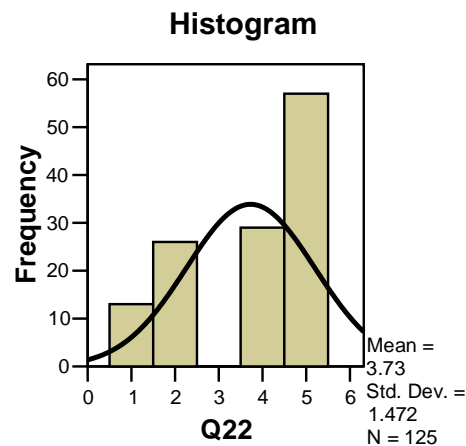
Responses to Q22 indicated that more than 74% students (M=3.89, STDV=1.319 Variance=1.741) and about 69% (M=3.73, STDV=1.472, Variance= 2.167) teachers were with the opinion that teachers used Bengali, their mother tongue, as a medium of instruction, that was; they used Bengali along with English while teaching English.. Q19 and Q22 were used to check the internal reliability and for cross referencing questions for each other. Q22 asked what methodology they believed was used in their instruction; more than 74% students and nearly 69% teachers reported that teachers' mode of instruction was a combined one. It provided

sufficient evidence that, they did not use the target language adequately in the class which directly opposed the principle of CLT.

Regardless of the respondents' accounts of the HSC examination impact, their responses to the above questions seem to indicate that the HSC examination in EFL has induced a certain degree of negative washback on their teaching practices in terms of time allocation, teaching focus, and teaching contents. However, to confirm the washback effects of the HSC examination, these responses need to be triangulated with other data sources (e.g. observations, interviews, and classroom observation). Q24 asked the respondents to assess their own pedagogical knowledge (e.g., whether they knew how to go about things in the course of their instruction and whether they were clear on the principles underpinning CLT). Correspondingly, Q6 and Q12 were designed to assess the respondents' actual understanding of CLT. The histograms (Figure 5.47 to Figure 5.50) present overall findings of this sub-section.

Figure 5.49: Teaching the meaning (student)
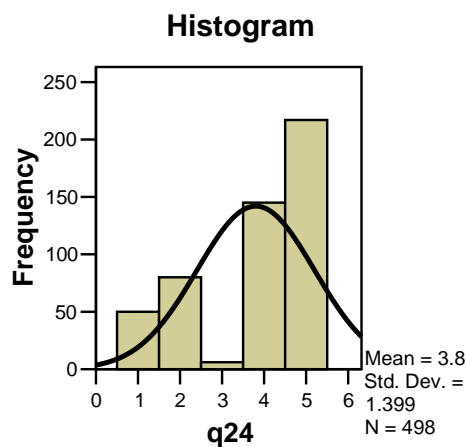
Figure 5.50: Teaching the meaning (teacher)

Histogram

Histogram

Mean = 3.8
Std. Dev. = 1.399
N = 498

Mean = 3.14
Std. Dev. = 1.494
N = 125
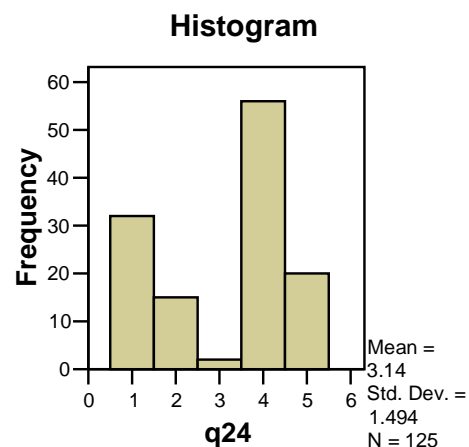
The results of Q24 revealed that over 72% (M=3.80, STDV=1.34) students and nearly 61% (M= 2.86, STDV=1.50) teachers demonstrated a poor understanding of the meaning of CLT; therefore, the teachers taught the meaning and theme of the topic and content of the textbook, *English for Today* (for classes 11-12). The findings indicated that the EFL teachers communicated the knowledge and meaning of the topic it contained as if they were teaching subjects such as history, or geography.

245

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

Language teachers should teach any lesson or topic from linguistic point of view. They must not put so much emphasis in providing subject matter and its inherent knowledge. Such evidence reveals that quite a number of teachers still have not achieved a good understanding of CLT or they do not understand CLT adequately.

5.1.4.2.3 Teachers' Encouragement and Motivation

Q20 and Q21 asked whether their teachers encouraged the students to speak English and ask question in the class. Over 61% students (M=2.66, STDV =1.53) suggested that their teacher did not encourage them to ask any question while more than 55% teachers (M=2.83, STDV =1.52) supported the students' view (Table 5.39). In replying to Q21, more than 54% students (M=2.83, STDV =1.51) commented that teachers did not motivate them to speak English, but nearly 53% teachers (M=3.19, STDV =1.53) claimed that they did encourage their students:

Table 5.39: Frequency counts on teachers' encouragement and motivation

| | Significant Frequency | | | | Negligible Frequency | | | | |
| | Agreement (SA+A) | | Disagreement (SD +D) | | Neutral | | Missing | | Total |
No	Freq.	Percent (%)	Freq.	Percent (%)	Freq	pct %	Freq	%	
SQ20	183	36.6	306	61.2	5	1.0	6	1.2	500
TQ20	52	41.6	69	55.2	1	.8	3	2.4	125
SQ21	222	44.4	271	54.2	2	.4	5	1.0	500
TQ21	66	52.8	56	44.8	3	2.4	-	-	125

Figure 5.51: Teachers' motivation (student)
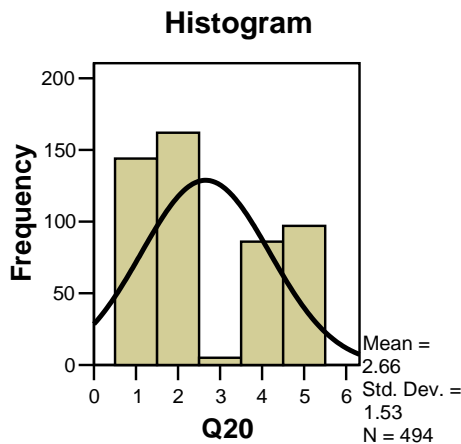
Figure 5.52: Teachers' motivation (teacher)

Histogram

Mean = 2.66
Std. Dev. = 1.53
N = 494

Histogram

Mean = 2.83
Std. Dev. = 1.52
N = 122
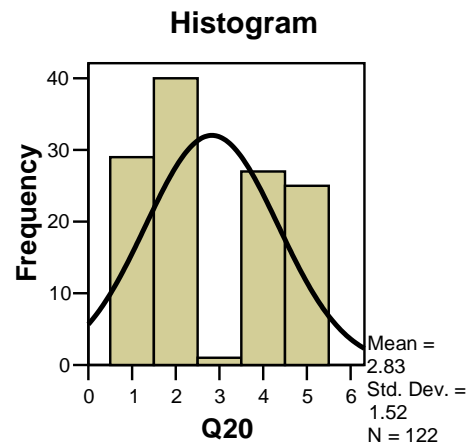
The variance for the Q20, and Q21 were respectively 2.34 and 2.30 for the student respondents; on the other hand, 2.30 and 2.35 for the teacher respondents. The medians for students were 2.00 for both questions while for teacher 2.00 and 4.00 for the questions Q20 and 21 respectively. These are normally distributed. The table (Table 5.40) below demonstrates the findings from descriptive statistics:

Table 5.40: Descriptive statistics on teacher's instructions of language

Qno	Mean	Median	Std. Deviation	Variance	Skewness	Kurtosis
SQ20	2.66	2.00	1.530	2.340	.432	-1.397
TQ20	2.83	2.00	1.520	2.309	.239	-1.528
SQ21	2.83	2.00	1.518	2.304	.167	-1.556
TQ21	3.19	4.00	1.533	2.350	-.138	-1.581

As shown in the figures (Figure 5.51 and Figure 5.52), the histograms for Q20 (Skewness=.432 for students, and skewness= .239 for teachers) are skewed positively which means both groups respondent disagreed with the statement of the question indicating teachers encouraged did not encouraged their students to asks questions in the class. In the case of Q21, the histogram of students is skewed positively while the teacher histogram for Q21 is skewed negatively (Figure 5.53 and Figure 5.54) which means that teachers disagreed with the students indicating that they encouraged their students to speak English. As shown in the histograms for both questions, the kurtosis values are properly distributed, and tails are proportionately shaped, indicating that the items appeared to be normally distributed.

Learners' active participation in the classroom activities is a precondition for an effective classroom. Learners' active participation in the language classroom is one of the principles of applying CLT. In Bangladesh, the classroom is still teacher-dominated where learners remain inactive as passive listeners. Wang (2008) shows that teacher factors influence teaching practices in the classroom. Teacher beliefs are consistent with their prior experience and instructional approaches. Watanabe (2000, p.44) notes that 'students rarely asked questions even during exam preparation lessons'. Cheng (1998) points out that while teachers talk less to the whole class as a result of the revised exam, the teacher talking to the whole class remains the dominant mode of interaction:

Figure 5.53: Encouragement and motivation (student)
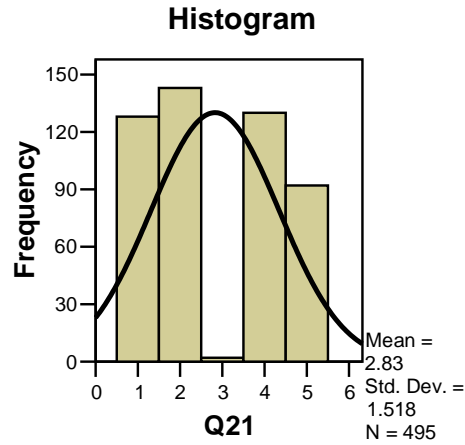
Figure 5.54: Encouragement and motivation (teacher)

Histogram

Histogram

Mean = 2.83
Std. Dev. = 1.518
N = 495

Mean = 3.19
Std. Dev. = 1.533
N = 125
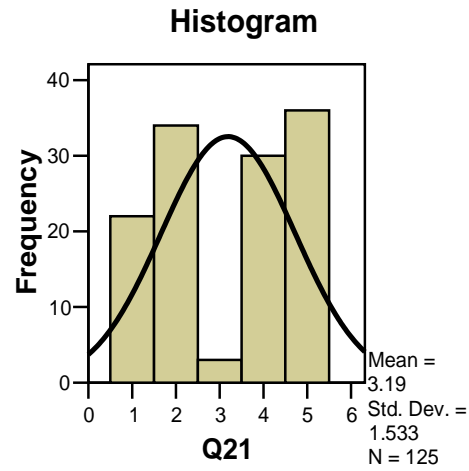
The concept, the 'teacher factor', has made its appearance in ELT in Bailey et al. (1996). In ELT and general education, there is a widely accepted assumption that teacher internal attributes such as encouragement, motivation, beliefs, assumptions, knowledge and experience make up the 'teacher factor' and this 'teacher factor' plays a powerful role both in determining teachers' perceptions of teaching and shaping their practices or actions in teaching (Richards, 2008)

5.1.4.2.4 Teaching to the Test

Q23 asked whether the teacher taught whatever he liked. In reply, over 70% students and over 71% teachers indicated that teacher taught whatever they preferred. It is now strongly grounded from the study that the dictates of high-stakes tests reduce the professional knowledge and status of teachers and exercise a great deal of pressure on them to improve test scores which eventually makes teachers experience negative feelings of shame, embarrassment, guilt, anxiety and anger:

Table 5.41: Frequency counts on teachers' teaching to the test

| | Significant Frequency | | | | Negligible Frequency | | | | |
| | Agreement (SA+A) | | Disagreement (SD +D) | | Neutral | | Missing | | Total |
No	Freq.	Percent (%)	Freq.	Percent (%)	Freq	pct %	Freq	pct %	
SQ23	351	70.2	141	28.2	7	1.4	1	.2	500
TQ23	89	71.2	34	27.2	2	1.6	-	-	125
SQ25	386	77.2	110	22	4	.8	-		500
TQ25	83	66.4	38	30.4	4	3.2			125

Vallette (1994) opines that washback is particularly strong in situations where the students' performance on a test determines future career options. In such case, teachers often feel obliged to teach to the test, especially if their effectiveness as a teacher is evaluated by how well their students perform:

Table 5.42: Descriptive statistics on teachers' teaching to the test

Qno	Mean	Median	Std. Deviation	Variance	Skewness	Kurtosis
SQ23	3.64	4.00	1.404	1.970	-.757	-.858
TQ23	3.74	4.00	1.442	2.079	-.836	-.805
SQ25	3.89	4.00	1.323	1.750	-1.066	-.198
TQ25	3.82	5.00	1.405	1.974	-.659	-1.225

In response to Q25, over 77% students (M= 3.89, STDV=1.323) pointed out that the teachers did not make them practise on how to learn and speak English, rather they taught how to answer the question to secure high score. Supporting the students' response, more than 66% teachers (M=3.82, STDV=1.405) suggested out that they taught how to prepare their students for the examination.

Swain (1985) says, "It has frequently been noted that teachers will teach to a test: that is, if they know the content of a test and/or the format of a test, they will teach their students accordingly" (p. 43). Tests are often perceived as exerting a conservative force which impedes progress. It is generally accepted that public examinations influence the attitudes, behavior, and motivation of teachers, learners and parents.

Figure 5.55: Teaching to the test (student)   Figure 5.56: Teaching to the test (teacher)
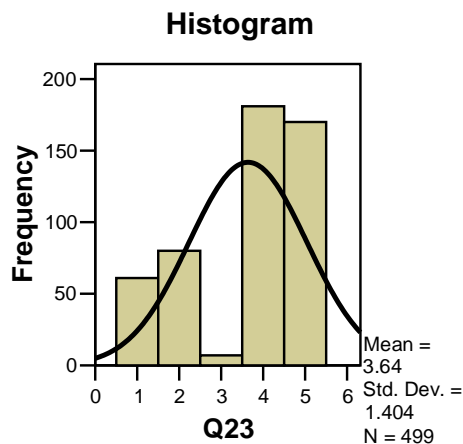
Histogram

Mean = 3.64
Std. Dev. = 1.404
N = 499

Histogram

Mean = 3.74
Std. Dev. = 1.442
N = 125

249

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>
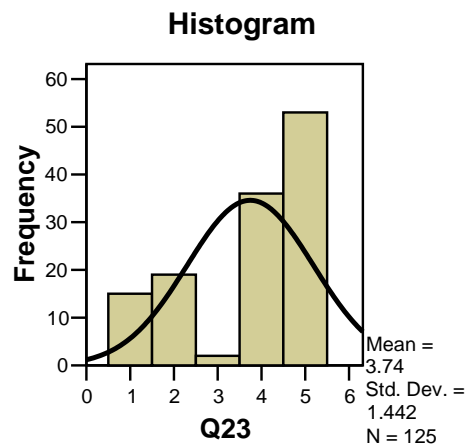
The tables (5.41 and 5.42) display the findings of the statistical analysis of the quantitative data. The mean scores, standard deviation, medians, variances, skewness and kurtosis values for each question are presented. The figures (Figure 5.55 to Figure 5.58) present the findings distributing the descriptive statistics along with skewness and kurtosis value. The histograms show the frequency, means, standard deviations, and valid sample size for each question. The main purpose of the presentation of histograms is to project the values of skewness and kurtosis of responses to the above questions:

Figure 5.57: Learning and speaking English (student)
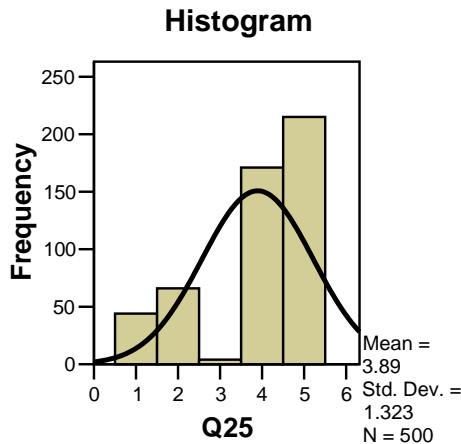
Figure 5.58: Learning and speaking English (teacher)
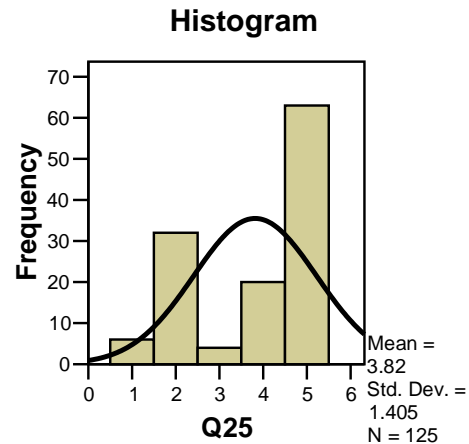
Empirical studies done by a number of researchers (e.g. Wang 2010) point out that the potential tests have to positively influence the methodology teachers usually use in classrooms, goes largely unrealized (Alderson and Wall, 1996; Watanabe, 1996; Nambiar and Ransirini: 2006). Communicative language tests, which include authentic test tasks, are fertile ground to be exploited in this regard. Theoretically, such tests have the potential for influencing teachers to step aside from the routine to create more innovative, student centered classrooms. As Hughes (2003) points out test impact is one of the most crucial considerations in communicate language teaching and testing.

5.1.4.2.5 Indication and Reflection of the HSC Examination Results

Question 26 (Q26) asked whether the participants believed that the test scores of the HSC examination in English were an appropriate indicator of a

250

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

student's English ability. The results revealed that over 63% students (M=2.64, STDV=1.488, Variance=2.214) believed that the HSC examination results would not indicate their language proficiency, whereas approximately 77% teachers (M=2.21, STDV=1.303, Variance= 1.698) agreed with the students indicating that HSC examination results would not reflect the students' language proficiency. The details of findings of this question are presented in the tables (Table 5.43 and Table 5.44) and histograms (Figure 5.59 and Figure 5.60):

Table 5.43: Frequency counts on teachers' teaching to the test

| | Significant Frequency | | | | Negligible Frequency | | | | |
| | Agreement (SA+A) | | Disagreement (SD +D) | | Neutral | | Missing | | Total |
Qno.	Freq.	Percent (%)	Freq.	Percent (%)	Freq	pct %	Freq	pct %	
SQ26	172	34.4	317	63.4	9	1.8	2	.4	500
TQ26	28	22.4	96	76.8	1	.8	-	-	125

As a cross-referencing question, Q6 asked whether the HSC examination tested student's proficiency in English. In reply to this question, nearly 70% students and about 60 % teachers believed that that the test was not valid and could not reflect actual teaching and learning or proficiency in English. Both the histograms are positively skewed indicating that respondents disagreed with the statement of the question. The values of kurtosis are distributed normally and frequently (Figure 5.59 and Figure 5.60):

Table 5.44: Descriptive statistics on teaching to the test

Qno	Mean	Median	Std. Deviation	Variance	Skewness	Kurtosis
SQ26	2.64	2.00	1.488	2.214	.507	-1.276
TQ26	2.21	2.00	1.303	1.698	1.027	-.180

The present study performed analyses HSC EFL test (English question papers) and answer scripts (section 5.53). While conducting analyses of HSC examination papers, and the HSC EFL answer scripts, it was found that the validity of the HSC examination in English, in respect of communicative language testing, was doubtful. The HSC examination did not test all the skills of English language

251

(e.g. speaking, listening). The reliability of the grading/scoring system was found faulty and motivated; the scoring unexpectedly varied from examiner to examiner to a large extent. The findings of the interview with the EFL examiners supported the findings of the questionnaire survey. Since the present HSC examination does not test listening and speaking skills at all, the examination results can not be an indicator of students' language proficiency:

Figure 5.59: Indicator of English language proficiency (student)
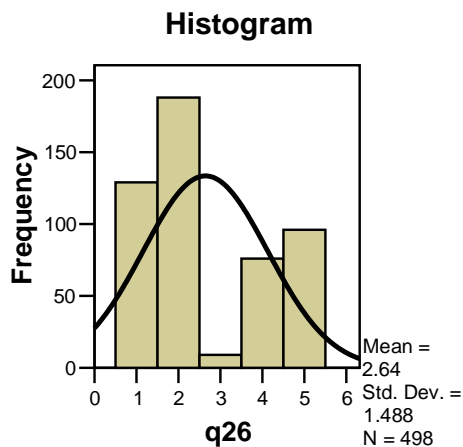
Figure 5.60: Indicator of English language proficiency (teacher)

Histogram

Mean = 2.64
Std. Dev. = 1.488
N = 498
q26

Histogram

Mean = 2.21
Std. Dev. = 1.303
N = 125
Q26
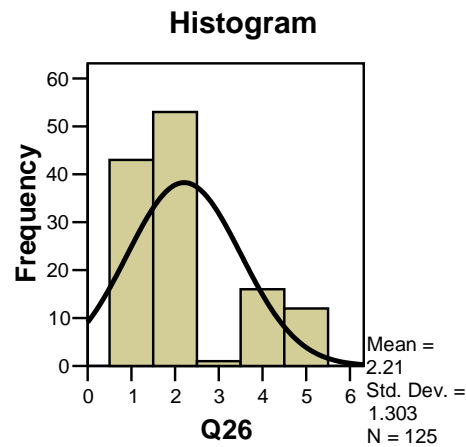
The findings are supported by Karabulut (2007) who finds that high school students and teachers focus more on the immediate goal of language learning which was to score high on the test and be admitted to the university by cramming for the test, and learning and practicing the language areas and skills that were measured on the test (grammar, reading, vocabulary) and ignored the ones that were not tested (listening, speaking, writing).

Pierce (1992) states "the washback effect, sometimes referred to as the systemic validity of a test (p.687). Cohen (1994) describes washback in terms of" how assessment instruments affect educational practices and beliefs" (p. 41). Pierce (1992, p.687) specifies classroom pedagogy, curriculum development, and educational policy as the areas where washback has an effect. On the other hand, Alderson and Hamp-Lyons (1996) take a view of washback which concentrated more on the effect of the test on teaching. They refer to washback as "… the influence that writers of language testing, syllabus design and language teaching believe a test will have on the teaching that precedes it" (ibid, p. 280). Bailey's

(1999) extensive summary of the current research on language testing washback highlights various perspectives, and provides deeper insight into the complexity of this phenomenon.

It is found that washback of the HSC examination in English exerts harmful influence on teaching methods and classroom instructions. Furthermore, it is clear from the study that it is the examination that generates less interaction in classes, or whether this is due to teachers believing, for whatever reason, that this is the way exams should be prepared for. The type and amount of washback on teaching methods appears to vary from context to context and teacher to teacher. It varies from no reported washback to considerable washback. The variable in these differences appears to be not so much the exam itself as the teacher. In this study, adequate evidences are found, which can determine that the washback influence the teaching methods and the ways they teach at the HSC level in Bangladesh context.

5.1.5 Classroom Tasks and Activities

The section discusses the findings of the 6 questions (Q27- Q32) related to the classroom tasks and activities, and their relation with washback effect of HSC public examination in English. The findings from frequency counts, mean score, standard deviation, skewness, kurtosis, etc for each question are discussed and presented by themes and issues. The skewness values, kurtosis, variances of the response are shown in histograms. Main issues presented and discussed here are: (a) tasks preferences (Q27 and Q29), (b) practise of model test (Q28 and Q30), and (c) examination pressure and teaching learning strategies (Q31 and Q32). The presentation, discussion, and reference proceed in coherent manner. The findings derived from inferential analyses of the questionnaire data are also presented.

5.1.5.1 Classroom Tasks and Activities Preferences

Question 27 and Q29 were asked to find out which activities were preferred by the teachers and students in the class. Q27 sought to find out whether the respondents had ever been exposed to task-oriented activities. The results suggested that nearly 78% students (M=3.95, STDV=1.317) and closely 62% teachers (M=3.43, STDV=1.552) immensely concentrated on task-oriented activities and

ignored the tasks and activities that are not directly related to passing the examination (Table 5.45 and Table 5.46):

Table 5.45: Frequency counts on tasks and activities preferences

| | Significant Frequency | | | | Negligible Frequency | | | | |
| | Agreement (SA+A) | | Disagreement (SD +D) | | Neutral | | Missing | | Total |
No	Freq.	Percent (%)	Freq.	Percent (%)	Freq	pct %	Freq	pct %	
SQ27	389	77.8	103	20.6	8	1.6	-	-	500
TQ27	77	61.6	47	37.6	1	.8	-	-	125
SQ29	353	70.6	141	28.2	5	1.0	1	.2	500
TQ29	85	68	38	30.4	2	1.6			125

There is a natural attitude for both teachers and students to tailor their classroom activities to the demands of the test, especially when the test is very important to the future of the students, and pass rates are used as a measure of teacher success.

In response to the question Q29, about 71% students (M=3.72, STDV=1.322, Variance= 1.748) replied that they spent more time practising grammar, and vocabulary related items because they (items) were tested in the examination. This view was supported by 68% teachers (M=3.58, STDV=1.466, Variance=2.148). The histograms below display skewness and kurtosis values along with other distribution of findings:

Figure 5.61: Ignoring tasks and activities (student)
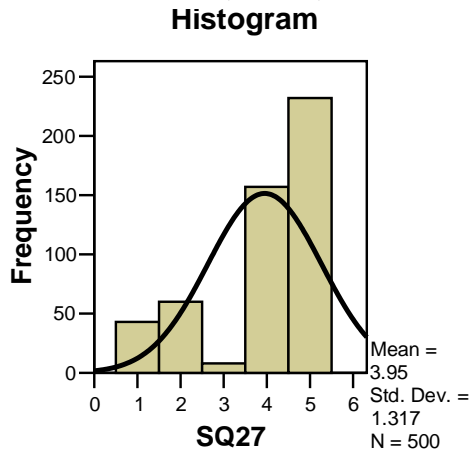
Figure 5.62: Ignoring tasks and activities (teacher)
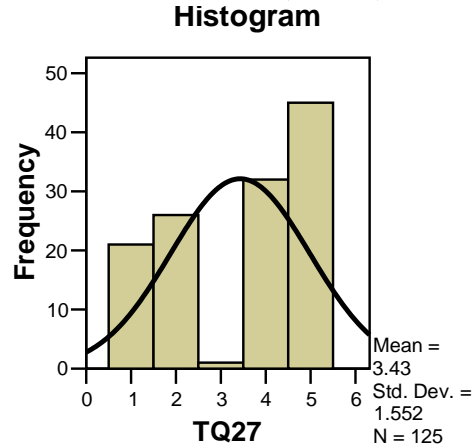
In recent years, though under researched, washback has become a much discussed topic among many linguistic and educational experts, and many of them admit that washback does exist and plays an important role in language teaching and learning. The histograms below (Figure 5.63 and Figure 5.64) show how the tails are skewed negatively and kurtosis values are distributed:

Figure 5.63: Practice of grammar and vocabulary items (student)
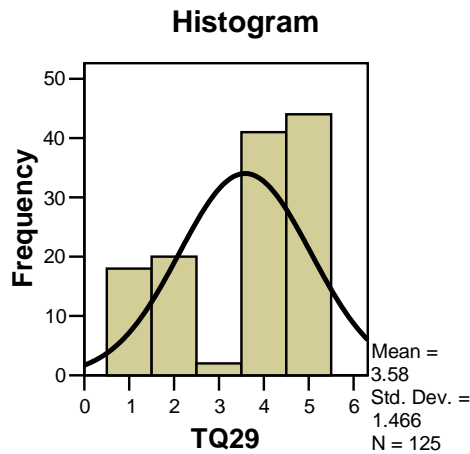
Figure 5.64: Practice of grammar and vocabulary items (teacher)
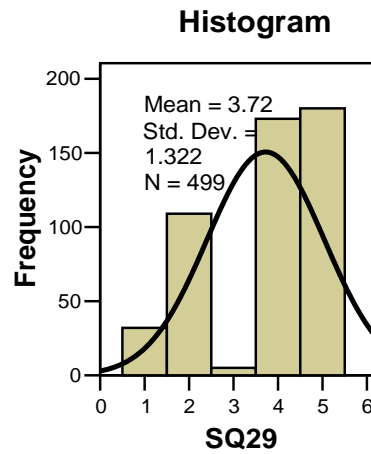
Table 5.46: Descriptive statistics on tasks and activities preferences

Qno	Mean	Median	Std. Deviation	Variance	Skewness	Kurtosis
SQ27	3.95	4.00	1.317	1.735	-1.133	-.045
TQ27	3.43	4.00	1.552	2.409	-.437	-1.441
SQ29	3.72	4.00	1.322	1.748	-.718	-.891
TQ29	3.58	4.00	1.466	2.148	-.684	-1.044

The findings of Q27 and Q29 are supported and cross-examined by the findings of classroom observation and interviews. While observing the teachers (during classroom observation), the present researcher found that most of the time they were teaching grammar, vocabulary, etc. The evidence of harmful washback is observed on classroom tasks and activities. The interviewed teacher stated that it was their responsibility to prepare their students for the forthcoming HSC examination. The findings are authenticated by the study of Lopez (2005). The researcher finds that there are matches and mismatches between the task and classroom practices. This influence of the test on the classroom (referred to as washback by language testers) is, of course, very important; this washback effect can be either beneficial or harmful (Buck, 1988).

5.1.5.2 Practice of Model Tests and Preparation Tests

Q28 asked whether teachers gave model test as a means of examination preparation. In response, about 80% students (M=3.95, STDV=1.252, Variance=1.567) replied that their teachers gave model examination before the final examination started. Again, more than 90% teachers (M=4.22, SRDV=.972, Variance=.945) confirmed that they used to give model tests so that their students could get familiar with the examination system and test contents (Table 5.47 and Table 5.48):

Table 5.47: Frequency counts on practice of model test and preparation test

| No | Significant Frequency | | | | Negligible Frequency | | | | Total |
| | Agreement (SA+A) | | Disagreement (SD +D) | | Neutral | | Missing | | |
	Freq.	Percent (%)	Freq.	Percent (%)	Freq	pct %	Freq	pct %	
Q28	398	79.6	95	19	7	1.4	-	-	500
Q28	113	90.4	12	9.6	-	-	-	-	125
Q30	411	82.2	83	16.6	4	.8	2	.4	500
Q30	102	81.6	22	17.6	1	.8			125

Washback is the power of examinations over what takes place in the classroom (Alderson and Wall, 1993). Swain succinctly suggests that it has frequently been noted that teachers will teach to a test: that is, if they know the content of a test and/or the format of a test, they will teach their students accordingly (Swain, 1985, p. 43). The figures below (Figure 5.65 to Figure 5.68) present the skewness and kurtosis values of Q28 and Q30:

Figure 5.65: Practice of model tests (student)
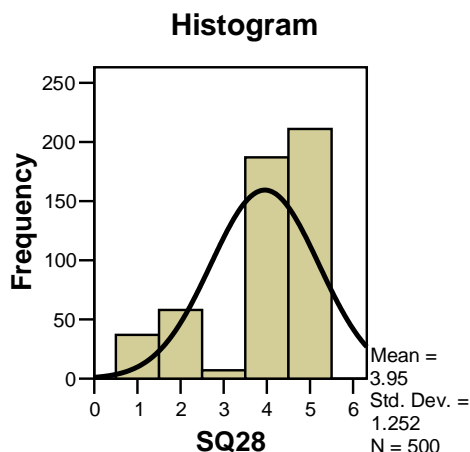
Figure 5.66: Practice of model tests (teacher)
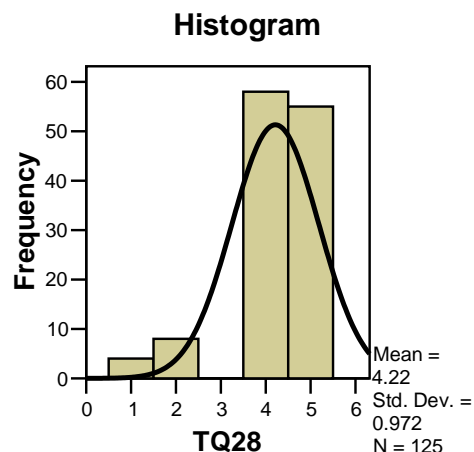
Responses to Q30, over 82% learners (M= 4.05, STDV= 1.181, Variance= 1.395) pointed that their teachers made them practise and solve the questions of past examinations so that they could be acquainted with test format and nature of items types; whereas nearly 82% teachers (M=4.00, STDV=1.270, Variance=1.613) disclosed that they made their students solve and practise the questions of past examination. Linguists often decry the 'negative' washback effects of examinations and regard washback as an impediment to educational reform or 'progressive' innovation in schools. Heyneman (1987) comments it's true that teachers teach to an examination. The table below presents the descriptive statistics of Q28 and Q30:

Table 5.48: Descriptive statistics on practice of model test and preparation test

Qno	Mean	Median	Std. Deviation.	Variance	Skewness	Kurtosis
SQ28	3.95	4.00	1.252	1.567	-1.181	-.045
TQ28	4.22	4.00	.972	.945	-1.733	3.010
SQ30	4.05	4.00	1.181	1.395	-1.311	.674
TQ30	4.00	4.00	1.270	1.613	-1.320	.583

Figure 5.67: Practice of past questions (student)
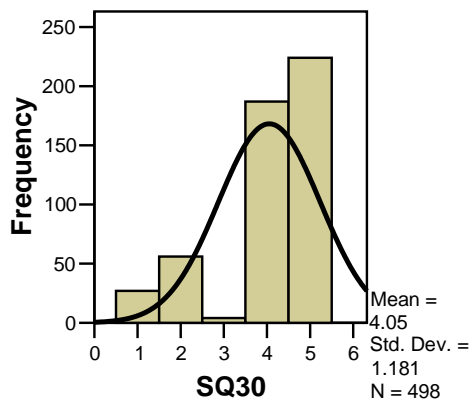
Figure 5.68: Practice past questions (teacher)
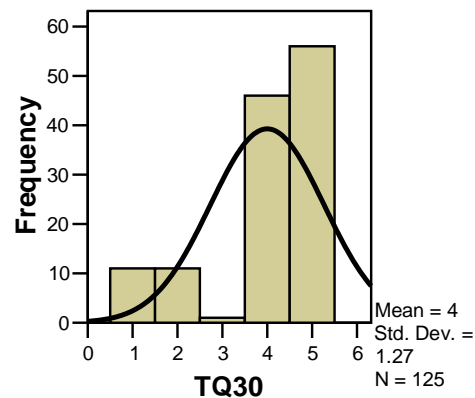
As the histograms (Figure 5.65 to Figure 5.67) have negative skewness (e.g. SQ28= -1.181, and TSQ28= -1.733) and therefore are negatively skewed; when the tails of histograms are negatively skewed, it indicates that most of the respondents agree with the statement. It was also found that the kurtosis values of the histograms were properly distributed as per the frequency of responses.

5.1.5.3 Examination Pressure and Teaching-Learning Strategies

Q31 asked the respondents whether the examination hindered and discouraged teaching and learning EFL. In reply to the question, nearly 70% students (M=3.70, STDV=1.388) replied that they gave little attention to learning English under test pressure and thus the HSC examination hindered their EFL learning while more than 70% (M=3.76, STDV=1.382) teachers supported the students' view on this issue indicating that test hindered their EFL teaching. The tables below (Table 5.49 and Table 5.50) present different levels of findings of this sub-section:

Table 5.49: Frequency counts on examination pressure and teaching learning

| | Significant Frequency | | | | Negligible Frequency | | | | |
| | Agreement (SA+A) | | Disagreement (SD +D) | | Neutral | | Missing | | Total |
No.	Freq.	Percent (%)	Freq.	Percent (%)	Freq	pct %	Freq	pct %	
Q31	349	69.8	144	28.8	5	1.0	2	.4	500
TQ31	88	70.4	35	28	2	1.6	-	-	125
SQ32	385	77	106	21.2	8	1.6	-	-	500
TQ32	103	82.4	20	16	2	1.6	-	-	125

The histograms (Figure 5.69 and Figure. 5.70) project the skewness, kurtosis, mean, standard deviation, and frequency of the findings (Q31). The histograms below show how they are skewed; the kurtosis values are distributed. The study found that the histograms were properly shaped with peaks and flats, and were normally distributed:

Table 5.50: Descriptive statistics on examination pressure and teaching learning

Qno	Mean	Median	Standard Deviation (STDV)	Variance	Skewness	Kurtosis
SQ31	3.70	4.00	1.388	1.928	-.734	-.923
TQ31	3.76	4.00	1.382	1.910	-.769	-.883
SQ32	3.96	4.00	1.328	1.764	-1.113	-.139
TQ32	4.12	5.00	1.182	1.397	-1.368	.797

Figure 5.69: Examination and
language learning (student)
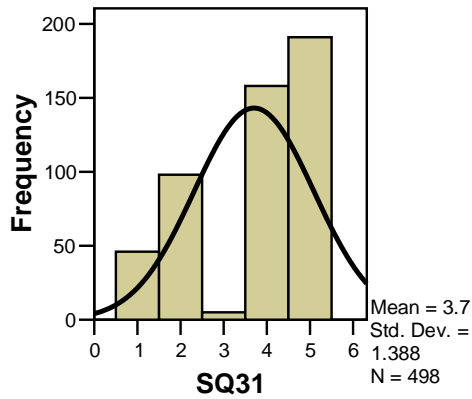
Figure 5.70: Examination and
language teaching (teacher)

Mean = 3.7
Std. Dev. =
1.388
N = 498

SQ31

Mean =
3.76
Std. Dev. =
1.382
N = 125
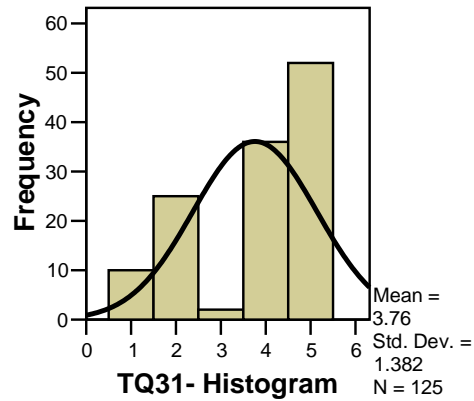
TQ31- Histogram

Q32 asked the respondents whether the teachers gave guidelines or taught
test taking strategies. As shown in the tables (Table 5.49 and Table 5.50) and
histograms (Figure 5.71 and Figure 5.72), 77% students (M=3.96, STDV= 1.328,
Variance=1.764) confirmed that the teachers taught them test taking guidelines and
strategies. Similarly, more than 82% teachers (M=4.12, STDV=1.182,
Variance=1.397) pointed that they really taught their students test taking strategies.
The result showed that the students were motivated more and spent more time in
preparing HSC examination. The findings are validated and cross-examined by the
findings of classroom observation and interview results. During the classroom
observation, the present researcher found that the EFL teachers spent a large amount
of time giving instructions on how to answer the questions in the examination:

Figure 5.71: Test- taking strategies
(student)
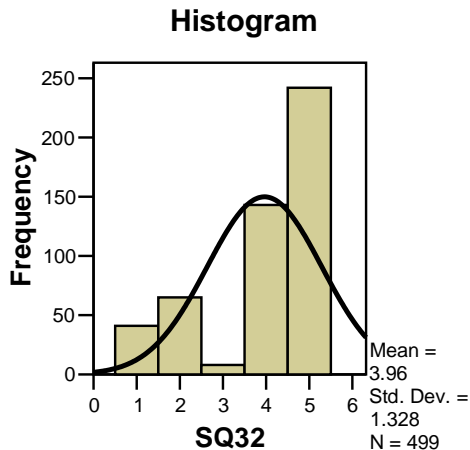
Figure 5.72: Test-taking strategies
(teacher)

Histogram

Histogram

Mean =
3.96
Std. Dev. =
1.328
N = 499

SQ32

Mean =
4.12
Std. Dev. =
1.182
N = 125

TQ32

259

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>
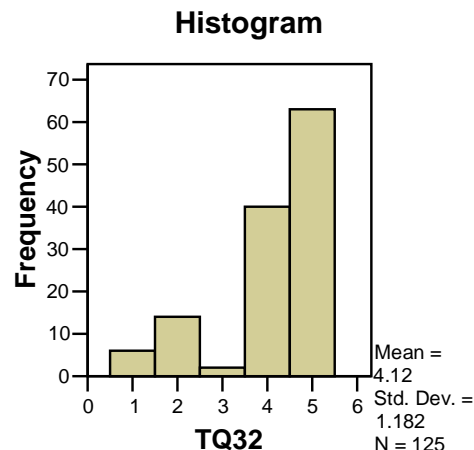
Test design and test-taking strategies are more closely identified with washback direction, while logistical issues are more closely identified with washback intensity (Kellaghan and Greene, 1992; Hughes, 1993). Through this study, it is adequately proved that the washback influences the test takers directly by affecting language learning (or non-learning), while the influences on other stakeholders affect efforts to promote language learning. The test-takers themselves can be affected by: the experience of taking and, in some cases, of preparing for the test; the feedback they receive about their performance on the test; and the decisions that may be made about them on the basis of the test.

5.1.6 Teaching of Language Skills and Elements

A group of questions (Q33- Q37) were asked to both students and teachers to know which skills were taught in the class. Question 33 (Q33) asked whether they practiced the EFL skills and elements as per the teacher's design and decision. More than 69% of the students (M= 3.73, STDV= 1.462, Variance=2.138) answered that their teacher designed the class activities himself. For the same question (Q33), more than 71% teachers (M= 3.83, STDV=1.474, Variance=2.173) agreed that they designed their class on their own decision. It was found that the class activities were teacher dominated; teacher taught whatever they liked. The teachers preferably taught those items which were related to test contents. The tables below (Table 5.51 and Table 5.52) present the details of findings in this section:

Table 5.51: Frequency counts on teaching of language skills and elements

| | Significant Frequency | | | | Negligible Frequency | | | | |
| | Agreement (SA+A) | | Disagreement (SD +D) | | Neutral | | Missing | | Total |
No.	Freq.	Percent (%)	Freq.	Percent (%)	Freq	pct %	Freq	pct %	
SQ33	346	69.2	141	28.2	10	2.0	3	.6	500
TQ33	89	71.2	36	28.8	-	-	-	-	125
SQ34	123	24.6	356	71.2	6	1.2	14	2.8	500
TQ34	21	16.8	104	83.2	-	-	-	-	125
SQ35	137	27.4	349	69.8	6	1.2	8	1.6	500
TQ35	40	32	84	67.2	1	.8	-	-	125
SQ36	380	76	106	21.2	6	1.2	8	1.6	500
TQ36	70	56	55	44	1	.8	2	1.6	125
SQ37	418	83.6	69	13.8	3	.6	10	2.0	500
TQ37	106	84.8	16	12.8	1	.8	2	1.6	125

Q34 was asked to ascertain if listening was practiced in the class. In response, over 71% students (M= 2.34, STDV=1.677, Variance= 2.814) replied that listening was not practised in the class, whereas 83% teachers (M=2.04, STDV=1.125, Variance=1.265) reported that they did not help students practise listening. Q35 probed whether speaking was practiced in the class. Nearly 70% students (M=2.30, STDV=1.539, Variance=2.144) reported that speaking was not practised in the class, while 67% teachers (M= 2.61, STDV=1.539, Variance= 2.369) confirmed that they did not help students practise speaking.

Table 5.52: Descriptive statistics on teaching of language skills and elements

Qno	Mean	Median	Std. Deviation	Variance	Skewness	Kurtosis
SQ33	3.73	4.00	1.462	2.138	-.780	-.934
TQ33	3.83	5.00	1.474	2.173	-.856	-.897
SQ34	2.34	2.00	1.677	2.814	3.827	37.567
TQ34	2.04	2.00	1.125	1.265	1.269	.802
SQ35	2.30	2.00	1.464	2.144	.809	-.872
TQ35	2.61	2.00	1.539	2.369	.616	-1.227
SQ36	4.00	5.00	1.316	1.731	-1.113	-.169
TQ36	3.38	4.00	1.463	2.142	-.175	-1.625
SQ37	4.17	4.50	1.121	1.258	-1.491	1.307
TQ37	4.23	5.00	1.085	1.177	-1.563	1.616

Q36 asked whether reading was practiced in the class. In replying to this question, more than 76% students (M=4.00, STDV=1.463, Variance=2.142) confirmed that it was practised. Furthermore, 56% teachers (M=3.38, STDV=1.121, Variance=1.258) suggested that they practised reading skills in the class. The histograms below (Figure 5.73 to Figure 5.76) display the findings of the questions mentioned along with the distribution of skewness and kurtosis values:

Figure 5.73: Practice of reading (student)
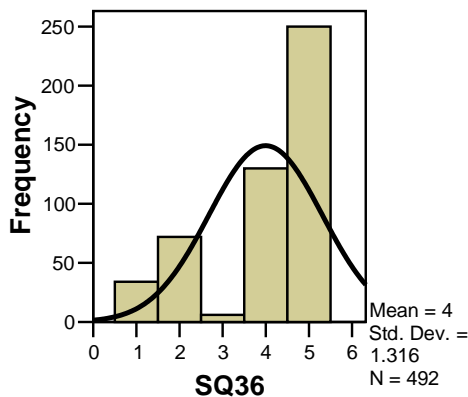
Figure 5.74: Practice of reading (teacher)
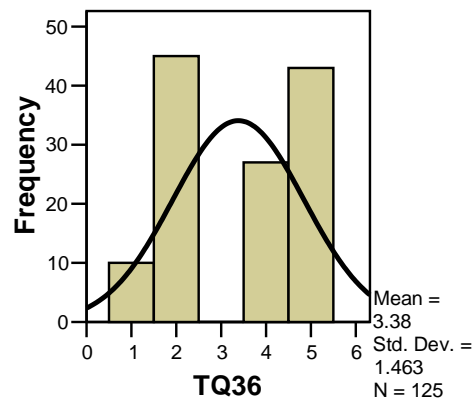
Q37 asked whether writing was practiced in the class. To reply to this question, approximately 84% students (M=4.17, STDV=1.121, Variance=1.258) commented that writing skills were practised, while almost 85% teachers (M=4.23, STDV=1.085, Variance=1.177) confirmed of practising the writing skills:

Figure 5.75: Practice of writing (student)
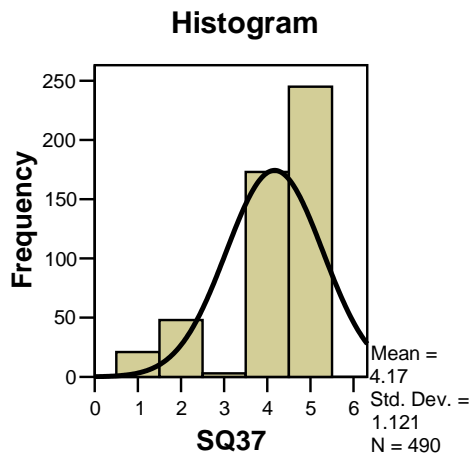
Figure 5.76: Practice of writing (teacher)

Histogram

Histogram

Mean = 4.17
Std. Dev. = 1.121
N = 490

Mean = 4.23
Std. Dev. = 1.085
N = 123
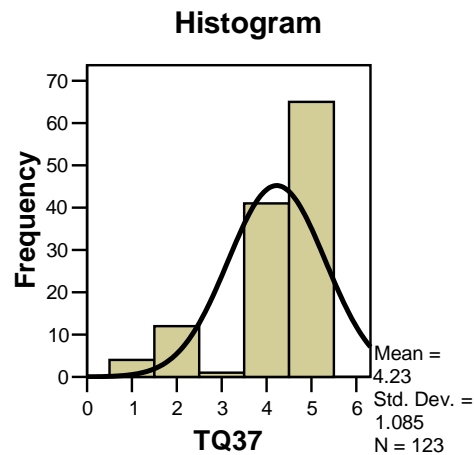
It was found that teachers taught those skills and elements (e.g. writing, reading, grammar, vocabulary, etc.) that were usually tested in the examination. There are some reasons for not practising two important skills: listening, and speaking. One of the reasons is that these skills (listening, and speaking) are not tested on the one hand, and they have little or no training in teaching the skills, on the other hand. The findings of this section were validated and cross-referenced with the findings of classroom observation and interview results.

During the classroom observation, the researcher found that most of the teachers (except T1) did not teach listening and speaking. The interviewed teachers also confessed that they avoided teaching of listening and speaking. Washback influences the teachers and test takers directly by affecting language. The teachers can affect both teaching and learning. The test-takers themselves can be affected by: the experience of taking and, in some cases, of preparing for the test; the feedback they receive about their performance on the test; and; the decisions that may be made about them on the basis of the test. Of the 15-washback hypotheses of Alderson and Wall's (1993, p. 120-121), five are directly address learner washback.

5.1.7 Beliefs, Attitudes and Perception as to the Test

Teachers' and students' beliefs in tests are likely to correspond to their beliefs in language teaching and learning. Their beliefs in language teaching and learning are likely to follow their conceptions of what is meant by learning as well as their beliefs in what language is. The relationship between beliefs in language teaching and beliefs in language learning is also interactive and interconnected. All these beliefs and attitudes are crucial in the sense that they may not only influence but also affect the way they interpret and react to washback. Such a basis not only helps to clarify the complexity of the innovation process, but also helps to improve further innovation endeavours.

5.1.7.1 The Descriptive Statistics

Washback may affect learners' actions and/or their perceptions, and such perceptions may have wide ranging consequences. This section dealt with 8 questions on different aspects. The internal reliabilities of the questions were: $\alpha =0.71$ (Cronbach's alpha) for student questions, and $\alpha= 0.77$ (Cronchbach's alpha) for teacher questions. The question addressed particular aspects of HSC EFL testing and its underlying influence on academic and personal behaviour. Major aspects addressed in this section were: (a) external and internal pressure for good results (Q38, Q44, and Q43), (b) anxiety and tension for examination (Q41), and (c) perception on HSC examination and its impact on future course of actions, (Q39, Q40, Q42, and Q45).

5.1.7.1.1 Perception of External Pressure and EFL Proficiency

Q38, Q44 and Q43 were asked to know about various issues on teachers' and students' perceptions on the HSC EFL test. Q 38 asked whether the respondents felt pressure for good results. More than 67% students (M=3.64, STDV= 1.530, Variance= 2.341) replied that their parents, college authorities, and relatives pressurised them to make good results. HSC examination is a high-stakes test and its result is of high importance for future career and education. Therefore, the parents and relatives feel concerned about their ward's results. In replying to the same question, 64% teachers (M=3.54, STDV=1.604, Variance=2.573) pointed out that

they felt external pressures (e.g. authority, guardians) to improve the pass rate and high scores in the examination. Very often, their reputation largely depends on the success rate in the examination. The tables below (Table 5.53 and Table 5.54) present the findings:

Table 5.53: Frequency counts on external and internal pressure and language proficiency

	Significant Frequency				Negligible Frequency				
	Agreement (SA+A)		Disagreement (SD +D)		Neutral		Missing		Total Sample
Qno.	Freq.	Percent (%)	Freq	Percent (%)	Freq	PCT %	Freq	PCT %	
SQ38	336	67.2	155	31	8	1.6	1	.2	500
TQ38	80	64	45	36	-	-	-	-	125
SQ43	400	80	93	18.6	6	1.2	1	.2	500
TQ43	104	83.2	19	15.2	1	.8	1	.8	125
SQ44	409	81.8	77	15.4	12	2.4	2	.4	500
TQ44	114	91.2	6	4.8	5	4.0	-	-	125

In response to Q43, 80% students (M= 4.05, STDV=1.278, Variance=1.634) believed that they could learn English better if there was no pressure for good results in the examination, while more than 82% teachers (M=4.00, STDV=1.176, Variance=1.382) pointed out that they could teach English better if there was no pressure. Pressure for good results impedes language teaching and learning; both external and internal pressure should be minimized   for creating a good teaching and learning atmosphere. The histograms below (Figure 5.77 and Figure 5.78) display different levels of findings and distribution of skewness and kurtosis:

Figure 5.77: Pressure for good results (student)
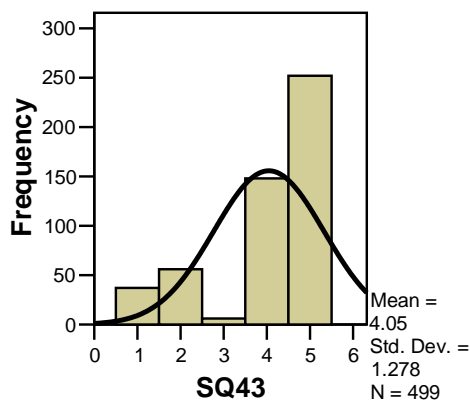
Figure 5.78: Pressure for good results (teacher)

Mean = 4.05
Std. Dev. = 1.278
N = 499

SQ43

Mean = 4
Std. Dev. = 1.176
N = 124

TQ43

264
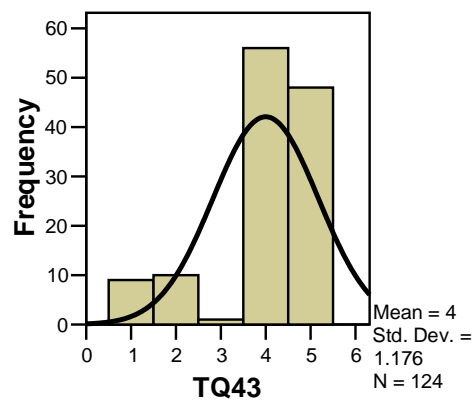
Q44 asked the respondents whether the students could make good results without improving their proficiency in English. Nearly 82% students (M=4.17. STDV=1.222, Variance= 1.493, skewness= -1.45, kurtosis=.912), and over 91% teachers (M=4.17, STDV=1.222, skewness=-1.912, kurtosis=4.547) believed that the learners could make their score relatively higher without attaining required level of proficiency in EFL. Test results have a significant impact on the career or life chances of individual test takers (e.g. educational/employment opportunities). They also impact on educational systems and on society more widely:

Table 5.54: Descriptive statistics on external and internal pressure and language proficiency

Qno	Mean	Mdn	STDV	Variance	Skewness	Kurtosis	95% Confidence Interval for Mean	
							Lower Bound	Upper Bound
SQ38	3.64	4.00	1.530	2.341	-.685	-1.142	3.51	3.77
TQ38	3.54	4.00	1.604	2.573	-.546	-1.411	3.26	3.82
SQ43	4.05	5.00	1.278	1.634	-1.267	.316	3.94	4.16
TQ43	4.00	4.00	1.176	1.382	-1.404	1.117	3.79	4.21
SQ44	4.17	5.00	1.222	1.493	-1.45	.912	4.06	4.28
TQ44	4.36	5.00	.865	.748	-1.912	4.547	4.21	4.51

The histograms below (Figure 5.79 and Figure 5.80) are negatively skewed indicating that most of the respondents agreed with the statements of the question. The kurtosis values in histograms are normally shaped, and the tails are normally distributed as per the frequency of the findings:

Figure 5.79: Language proficiency versus good results (student)
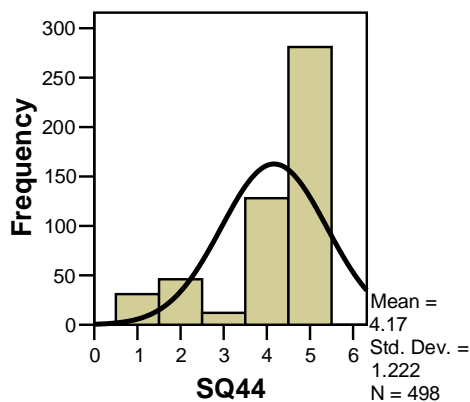
Figure 5.80: Language proficiency versus good results (teacher)

Mean = 4.17
Std. Dev. = 1.222
N = 498

SQ44

Mean = 4.36
Std. Dev. = 0.865
N = 125

TQ44
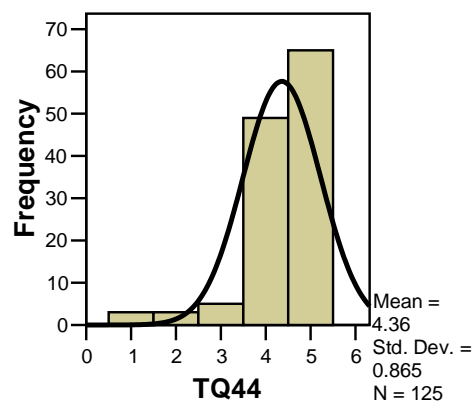
5.1.7.1.2 Anxiety and Tension for Examination

Q41 asked whether the students suffered from anxiety and tension for the HSC examination in English. 79% students (M=3.99, STDV=1.236, Variance=1.528) confirmed that they suffered from tension and anxiety, whereas more than 78% teachers (M=3.92, STDV= 1.323, Variance=1.752) supported the students' reply (Table 5.55 and Table 5.56):

Table 5.55: Frequency counts on anxiety and tension for examination

	Significant Frequency				Negligible Frequency				
	Agreement (SA+A)		Disagreement (SD +D)		Neutral		Missing		Total Sample
No.	Freq.	Percent (%)	Freq	Percent (%)	Freq	PCT %	Freq	PCT %	
SQ41	395	79	98	19.6	4	.8	3	.6	500
TQ41	98	78.4	24	19.2	3	2.4	-	-	125

Since the HSC examination is a high-stakes test, its impact on learners and teachers are manifold. Q38 was used as a cross-referencing question for this item where it was found that the both groups of respondents suffered from external pressures and those pressures generated anxiety and tension. The teachers as well as the learners are adequately aware of the consequences of failure or poor results in the examinations:

Table 5.56: Descriptive statistics on anxiety and tension for examination

Qno	Mean	Mdn	STDV	Variance	Skewness	Kurtosis	95% Confidence Interval for Mean	
							Lower Bound	Upper Bound
SQ41	3.99	4.00	1.236	1.528	-1.147	.094	3.88	4.10
TQ41	3.92	4.00	1.323	1.752	-1.187	.154	3.69	4.15

5.1.7.1.3 Perception of the HSC Examination in English

Q39, Q40, Q42 and Q43 asked the respondents how they felt with the EFL examination in English. Q39 asked whether the students could make high score without improving their language proficiency. For this, 72% teacher (M=3.69, STDV=1.431, Variance=2.048) replied in the affirmative, while nearly 74% teachers (M=3.92, STDV=1.423, Variance =2.026) agreed with the students (Table 5.57):

266

Table 5.57: Frequency counts on perception and belief

| | Significant Frequency | | | | Negligible Frequency | | | | |
| | Agreement (SA+A) | | Disagreement (SD +D) | | Neutral | | Missing | | Total Sample |
No.	Freq.	Percent (%)	Freq	Percent (%)	Freq	PCT %	Freq	PCT %	
SQ39	360	72	135	27	3	.6	2	.4	500
TQ39	92	73.6	31	24.8	2	1.6	-	-	125
SQ40	255	51	236	47.2	7	1.4	2	.4	500
TQ40	101	80.8	22	17.6	1	.8	1	.8	125
SQ42	127	25.4	365	73	5	1.0	3	.6	500
TQ42	34	27.2	88	70.4	3	2.4	-	-	125
SQ45	423	84.6	65	13	10	2.0	2	.4	500
TQ45	104	83.2	18	14.4	3	2.4	-	-	125

Washback may affect learners' actions and/or their perceptions, and such perceptions may have wide ranging consequences. Sturman (1996) uses a combination of qualitative and quantitative data to investigate students' reactions to registration and placement procedures at two English language schools in Japan. The placement procedures included a written test and an interview. He finds that students' perceptions and beliefs towards test contribute to the learning in school and at home.

In replying to Q40, asked 51% students (M=3.16, STDV=1.527, Skewness= -.065, Kurtosis= -1.6) considered the examination results as the feedback of their learning while nearly 81% teachers (M= 4.10, STDV= 1.174) believed that they got feedback of their teaching from the students' results (Table 5.57 and Table 5.58):

Table 5.58: Findings from descriptive statistics on perception and belief

| Qno | Mean | Mdn | STDV | Variance | Skewness | Kurtosis | 95% Confidence Interval for Mean | |
							Lower Bound	Upper Bound
SQ39	3.69	4.00	1.431	2.048	-.838	-.771	3.56	3.82
TQ39	3.92	5.00	1.423	2.026	-1.000	-.554	3.67	4.17
SQ40	3.16	4.00	1.527	2.332	-.065	-1.599	3.06	3.32
TQ40	4.10	4.50	1.174	1.379	-1.247	.365	3.89	4.31
SQ42	2.30	2.00	1.383	1.913	.884	-.608	2.18	2.42
TQ42	2.34	2.00	1.449	2.098	.802	-.830	2.09	2.59
SQ45	4.28	5.00	1.120	1.254	-1.611	1.549	4.18	4.38
TQ45	4.14	4.00	1.134	1.286	-1.401	1.032	3.94	4.34

Q42 asked whether the present HSC examination in English helped the students improve language proficiency. In reply, 73% students (M=2.30,

267

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

STDV=1.383, Variance=1.913) replied in the negative saying that the present examination system did not help them improve language proficiency. Similarly, almost 74% teacher (M=2.34, STDV=1.449, Variance=2.098) agreed with the maximum students. Q44 was used as a cross–referencing question for checking the validation of the results. As an another cross-referencing question, Q6 was used which asked whether the examination tested student's overall competence in English; it was found that more than 69% students and over 59% teachers replied that the overall competence in English was not tested:

Figure 5.81: Feeling embarrassed (student)
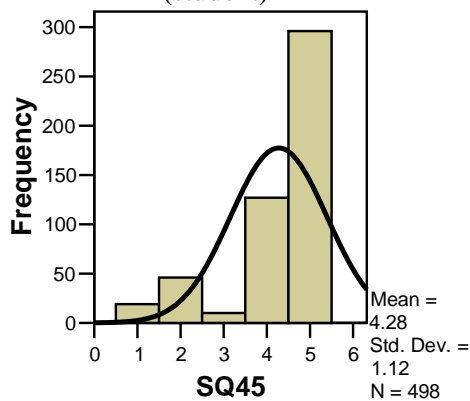
Figure 5.82: Feeling embarrassed (teacher)
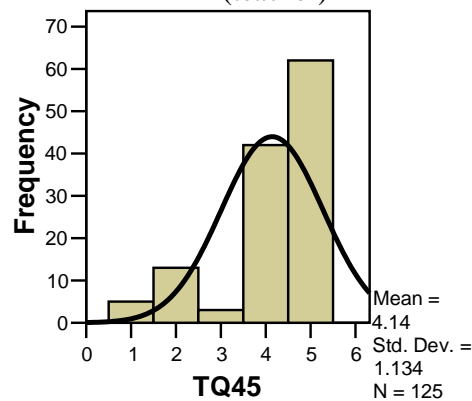
Q45 asked whether the respondents were frustrated or embarrassed incase of failure or poor performance in the examination. Almost 85% students (M=4.28, STDV=1.120, Variance=1.254) replied that they were frustrated if they failed or performed badly in the examination. Over 83% teachers (M= 4.14, STDV=1.134, Variance=1.286) commented that they were embarrassed about failure and poor performance in the examination. The histograms are skewed negatively (Figure 5.81 and Figure 5.82). The tables (Table 5.57 and Table 5.58) present the details of findings. Along with the mean and the standard deviation the histograms display the distribution of skewness and kurtosis values. A large number of teachers help students cope with the examinations in order to preserve their reputation as good teachers. This situation is unavoidable because of the extrinsic values of examinations (Khaniya, 1990).

Herman and Golan (1991 and 1993) indicated that teachers in schools with increasing test scores felt more pressure to improve their students' test scores from different external sources than teachers in schools with stable or decreasing scores did. The external sources included their principals, school administrators, other

268

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

teacher colleagues, parents, the community, and/or the media. In this study, the external forces, which existed within society, education and colleges, that influenced teachers' curricular planning and instruction, were examined. Teachers' perceived external pressure in teaching was measured by summating the total score of the items related to this domain on the survey questionnaire.

Linguists and EFL practitioners worldwide are now raising their voice for "testing for teaching, not teaching for testing". Test should be used as a lever of promoting learning. But in many countries like Bangladesh, due to adoption of poor education policy, the test itself hinders learning, especially, learning English as a foreign language. Tests can aid learning and teaching both if aimed to assess the required skills. It is now accepted that public examinations influence the attitudes, behaviour, and motivation of teachers, learners and parents. Many studies have been carried out on washback explicating that it can be either beneficial or harmful depending upon the contents and techniques.

The test is compulsory a part of education. All classes have tests, and all students are expected to perform to the best of their abilities on tests. Therefore, teachers and students place significant emphasis on tests despite the stakes. Andrews and Fullilove point out, "Not only have many tests failed to change, but they have continued to exert a powerful negative washback effect on teaching" (Andrews and Fullilove, 1994, p. 57). Tests are often perceived as exerting a conservative force which impedes progress. Heyneman (1987) has commented that teachers teach to an examination. In Bangladesh context, it is a proven fact that that the pass rate in the examination is the only measure to assess institutional success. It is a very common phenomenon that many candidates commit suicide and get absconded due o failure or poor performance in the examination. For many institutions, the salary of the teachers remains held up for poor success rate in the examination.

5.1.7.2 Levene's Test and T-Test Analysis

As mentioned, the independent-samples T-Test evaluates the difference between the means of two independent or unrelated groups. That is, it evaluates whether the means for two independent groups are significantly different from each other. The independent-samples T-Test is commonly referred to as a between-groups

design, and can also be used to analyze a control and experimental group. With independent-samples T-Test, each case must have scores on two variables, the grouping (independent) variable and the test (dependent) variable. The grouping variable divides cases into two mutually exclusive groups or categories, here, students and teachers. The T-Test evaluates whether the mean value of the test variable (e.g., test performance) for one group (e.g., students) differs significantly from the mean value of the test variable for the second group (e.g., teachers):

Table 5.59: Statistics on belief, attitudes and perception towards the test

	Resp_type	N	Mean	Std. Deviation	Std. Error Mean
Q38	Student	499	3.64	1.530	.068
	Teacher	125	3.54	1.604	.143
Q39	Student	498	3.69	1.430	.064
	Teacher	125	3.92	1.423	.127
Q40	Student	498	3.16	1.529	.069
	Teacher	124	4.10	1.174	.105
Q41	Student	497	3.99	1.236	.055
	Teacher	125	3.92	1.323	.118
Q42	Student	497	2.30	1.378	.062
	Teacher	125	2.34	1.449	.130
Q43	Student	499	4.05	1.278	.057
	Teacher	124	4.00	1.176	.106
Q44	Student	498	4.17	1.222	.055
	Teacher	125	4.36	.865	.077
Q45	Student	498	4.28	1.120	.050
	Teacher	125	4.14	1.134	.101

In this section, the findings of Levene's test for equality of variances and independent sample test (T-Test) are presented. This section includes 8 questions (Q38-Q45) about different issues on belief, attitudes and perception of the teachers and the students towards the HSC examination in English and its impact on their academic and personal behaviours.

In the present study, the researcher used T-Test (the Independent Samples T-Test) to compares the mean scores of students and teachers on a given variable. As mentioned, the first step in using the independent-samples T-Test statistical analysis is to test the assumption of homogeneity of variance, where the null hypothesis

270

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

assumes no difference between the two group's variances $(H_0: \sigma_1^2 = \sigma_2^2)$. The Levene's F Test for Equality of Variances is the most commonly used statistic to test the assumption of homogeneity of variance. The Levene's test uses the level of significance set a priority for the T-Test analysis.

- There is not a significant difference if the sig. value is greater than alpha(.050)
- There is a significant difference if the sig. value is less than or equal to alpha (.05)

    (e.g., $\alpha = .05$) to test the assumption of homogeneity of variance.

Table 5.60: Levene's test of equity of variances- significant deference

		Levene's Test for Equality of Variances	
		F	Sig.
Q40	Equal variances assumed	91.471	.000*
	Equal variances not assumed		

* = significant at p < 0.05

Table 5.61: T-Test for equity of means for significant difference

		Levene's Test for Equality of Variances		t-test for Equality of Means				95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Q40	Equal variances assumed	91.471	.000	-6.406	620	.000*	-.942	-1.231	-.653
	Equal variances not assumed			-7.492	238.23	.000*	-.942	-1.190	-.694

* = significant at p < 0.05

For Q40 (shown above), the F value for Levene's test was 91.471 with a Sig. (*p*) value of .000 ($p < .001$). Since the Sig. value was less than alpha of .05 ($p<.05$), the null hypothesis was rejected (no difference) for the assumption of homogeneity of variance. There was a significant difference between the two groups' variances (students and Teachers). That is, the assumption of homogeneity of variance was not met. If the assumption of homogeneity of variance is not met, the data results

associated with the "Equal variances not assumed," is taken. If the assumption of homogeneity of variance is met, the data results associated with the "Equal variances assumed," is taken and interpreted the data accordingly Table 5.60 and Table 5.61).

For this example (Q42), since the t value (--7.492, which indicates that the second group was higher than the first group) resulted in a Sig. (*p*) value that was (.000) less than the alpha of .05 (*p* < .05, which puts the obtained *t* in the tail) – the study rejected the null hypothesis in support of the alternative hypothesis, and concluded that students and teachers differed significantly on the same variable.

By examining the group means for this sample of subjects (not shown here), the study found that the teachers (with a mean of 4.10) responded significantly higher on this domain than students did (with a mean of 3.16). Similarly, for the Q43 (table: 5.71), the F value for Levene's test was 4.950 with a Sig. (*p*) value of .026 (*p* < .001). Because the Sig. value was less than alpha of .05 (*p* < .05), the study rejected the null hypothesis (no difference) for the assumption of homogeneity of variance and concluded that there was a significant difference between the two groups' variances (students and Teachers). Therefore, the bottom line data was used for T-Test results on variances. It was found that the t value= .384, and significance sig (2-tailed) = .702. The p value was greater that alpha value p>.05 which indicated that the variances were not significantly different.

For Q44, F value for Levene's test was 13.521 with a Sig. (*p*) value of .000 (*p* < .001). Since the Sig. value was less than alpha of .05 (*p* < .05), the study rejected the null hypothesis (no difference) for the assumption of homogeneity of variance and concludes that there is a significant difference between the two group's variances (students and Teachers). The Levene's test uses the level of significance set *a priority* for the *t* test analysis. Now the bottom line data "Equal variances not assumed" is used for T-Test analysis. It was found that t value = -2.019, df=262.796, and the sig (2-tailed) =.045. The significance of t (p-value) was smaller than the alpha value p<.05 which indicated that the variances between the two groups (students and teachers) were significantly different:

Table 5.62: Levene's test of equity of variances- insignificant deference

		Levene's Test for Equality of Variances	
		F	Sig.
Q38	Equal variances assumed	2.652	.104
	Equal variances not assumed		

* = significant at p < 0.05

For the Q38, the F value for Levene's test was 2.652 with a Sig. (*p*) value of .104 (*p* < .001). As the Sig. value was more than alpha of .05 (*p* > .05), the study accepted the null hypothesis (no difference) for the assumption of homogeneity of variance and concluded that the two groups were not significantly different (students and Teachers). That is, the assumption of homogeneity of variance was met (Table 5.62):

Table 5.63: T-Test for equity of means for insignificant difference

		Levene's Test for Equality of Variances		T-Test for Equality of Means				95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Q38	Equal variances assumed	2.652	.104	.591	622	.555	.091	-.212	.395
	Equal variances not assumed			.574	184.574	.567	.091	-.222	.405

* = significant at p < 0.05

If the Levene's Test is significant (the value under "Sig." is less than .05), the two variances are significantly different. If it is not significant (Sig. is greater than .05), the two variances are not significantly different; that is, the two variances are approximately equal (Table 5.63). If the Levene's test is not significant, the study has met the second assumption. Here, it was found that the significance was .104, which is greater than .05. We can assume that the variances are approximately equal. The students mean and teachers mean (Students Mean=3.64, and Teachers

273

Mean=3.54) are nearly equal. Therefore, the results associated with the "Equal

variances assumed was taken (Top line)" and interpreted the findings accordingly.

For Q38, it was found that t value=. 591, df= 622, and sig (2-tailed) =.555

which was greater than alpha value (p) .05. Therefore, it concluded that the two

variances were not significantly different. That is, the both group of respondents

gave similar opinion on the issue that the teachers took care of their students

whether the students could follow their teacher's instruction:

Table 5.64: Finding of T-Tests analysis: Independent Samples Test

		Levene's Test for Equality of Variances		T-Test for Equality of Means					
								95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Q38	Equal variances assumed	2.652	.104	.591	622	.555	.091	-.212	.395
	Equal variances not assumed			.574	184.574	.567	.091	-.222	.405
Q39	Equal variances assumed	.066	.797	-1.604	621	.109	-.229	-.510	.051
	Equal variances not assumed			-1.608	191.730	.109	-.229	-.510	.052
Q40	Equal variances assumed	91.471	.000	-6.406	620	.000*	-.942	-1.231	-.653
	Equal variances not assumed			-7.492	238.231	.000*	-.942	-1.190	-.694
Q41	Equal variances assumed	.834	.361	.541	620	.588	.068	-.179	.314
	Equal variances not assumed			.520	182.191	.604	.068	-.190	.326
Q42	Equal variances assumed	1.489	.223	-.346	620	.729	-.048	-.322	.225
	Equal variances not assumed			-.336	184.454	.737	-.048	-.331	.235
Q43	Equal variances assumed	4.950	.026	.365	621	.715	.046	-.202	.294
	Equal variances not assumed			.384	201.611	.702	.046	-.191	.283
Q44	Equal variances assumed	13.521	.000	-1.650	621	.100	-.191	-.419	.036
	Equal variances not assumed			-2.019	262.796	.045*	-.191	-.378	-.005
Q45	Equal variances assumed	.030	.862	1.167	621	.244	.131	-.089	.352
	Equal variances not assumed			1.159	189.331	.248	.131	-.092	.354

* = significant at p < 0.05

The Levene's (F) values for the Q39, Q41, Q42, and Q45 were .066, .834,

1.489, and .030 respectively. The Levene's significances (sig) of all those questions

274

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

were (P>50) .797.361, .223, and .862. The significances were greater than .05 which indicated that their variances were insignificantly different. Therefore, the 'Equal variances assumed' (top lines) output was used for T-Test analysis. For (variable) Q39, the t value was 1.604, df=621, and mot importantly the sig (2-tailed) was .109. This indicated that the variances between two groups were insignificantly different or nearly equal. Q39 asked whether the students could score good marks without improving their English proficiency. Levene's test for Equality of Variances as well as t-test proved that the both group of respondents gave almost same responses (students mean=.3.69, teachers mean=3.92) and agreed with statement of questions (Table 5.64).

The t value for Q41 was .541; df was 620; and the significance (2-tailed) was .588. The significance of t was insignificant because the p value was (p>.05) greater than alpha value .05. So, the variances of the two groups were not significantly different. Q41 asked whether the students suffered from anxiety and tension. It was found that both groups of respondents confirmed that the students suffered from anxiety and tension for the examination. For Q42, t value=.384, df=201.611, means difference =-.048, and significance of t= .702.

Levene's test for Equality of Variances and independent sample test (t-test) find that the means variances between two groups were not significantly different. The students' mean (2.30) and teachers' mean (2.34) were nearly same. The respondents believed that the present HSC examination in English did not help them improve language proficiency. The Levene's F value for Q45 was .030, and significance was .862 (sig). Levene's significance (P>.05) was greater than .05 which suggested that variances of two groups were negligible or insignificant. Furthermore, the t- value for this question was 1.167 and df was 621. Most importantly, the significance (2-tailed) was .244 which was larger than alpha value .05. Therefore, it concluded that the variances were not significantly different indicating that both groups of respondents equally agreed to the issue of embarrassment and frustration due to failure or poor performance in the examination.

5.1.8 Evidence of Washback from the Questionnaire Surveys

It can be seen from the findings of the questionnaire surveys that identifying washback effects of the HSC examination is complicated. It was found that very often students did not have much control over the choice of learning activities in the classroom. The activities they actually carried out in the class were assigned by their teachers. The teachers did it because of power of the test. Actually, the teachers taught to the test. It was a strong evidence of negative washback on teaching and learning. Differences in the teachers' and students' opinions and perceptions were tested for statistical significance using independent sample T-Tests. A probability of less than 0.05 was taken as statistically significant for both tests. In most of the cases, the difference between the teachers' and students' opinion was statistically insignificant indicating that they gave almost similar responses. There was much of an indication of negative washback on aspects of teaching at the micro level.

The study revealed that both the teachers and the students did not bother and care about the objectives of the HSC syllabus and curriculum; therefore, teaching and learning of communicative competence (e.g. listening, speaking, reading, and writing) were neglected in the class. The both groups of respondents gave a higher weighting to the activities of the test preparation in the classes, putting more emphasis on reading and writing. However, the teachers considered it less likely that they would employ new teaching methods to increase communicative competence, indicating a degree of reluctance by teachers to make changes in certain aspects of their teaching. In addition, the majority of the teachers adequately used Bengali (half English and half Bengali) as a medium of instruction due to, according to the teachers, the low level of their students' language proficiency. The study found that the examination influenced how the teachers would teach. The majority of the teachers employed test-oriented commercially produced materials and their teaching mainly relied on the content and organisation of the HSC examination. The results suggested that there was washback effect on the teaching materials. Most of the teachers and students changed their emotion, attitudes, and classroom behaviours because of the influence of the test. The results indicated that where there was a test impact, it was likely to be negative. Explaining the teachers' reluctance to make changes is complicated, and is explored further through detailed classroom observations, review of examination related documents and in-depth interviews in the next sections of this chapter.

5.2 Findings of the Classroom Observation

For the presentation and discussion of the findings from the classroom observation, the present researcher initially gives a description of the background information of the observation schedules and checklist. Then, he addresses personal details of the participants including their gender, education level completed, number of years of teaching, previous language teaching experiences, and training received in ELT. Following that, the researcher reports their (teachers') teaching performances which he examined in their classrooms.

In light of the abundance of data collected, only general information is given here. Specifically, the present researcher focuses only on the three major themes that have emerged pertaining to the research questions. The first theme describes the influence of the HSC examination in English on teaching and learning. It includes beliefs teachers articulated in the HSC examination and its impact (washback). The second theme reports on teachers' curriculum knowledge (e.g., knowledge of the HSC examination and the Syllabus and Curriculum). The third theme presents the real evidence of the effects of the HSC examination on their teaching that has emerged during the classroom observation. The last theme has emerged under the category of teachers' beliefs in teaching and classroom scenarios of how they teach.

It describes teachers' various conceptions of teaching and learning and their real practice in the classroom. In terms of teacher practice, the classroom scenarios portrayed involves their interaction patterns, various activities organised, focus of instruction (e.g., focus on knowledge or competence), skills practiced, materials used , personal behaviour and characteristics, and medium of instruction. The presentation and discussion of findings derived from different schedules and checklist are made one by one. First, the researcher presents and discusses the findings collected by COLT. Then, he reports the findings obtained from the use of UCOS. Finally, the present researcher offers the findings resulted from a self-made observation checklist.

5.2.1 Observation Schedules and Checklist

As stated in Chapter Four, a number of observation instruments have been applied based on developments in language teaching. However, no observation instruments have been developed specifically for washback studies. As a result, the classroom observation instruments were designed for the purpose of the present study in Bangladesh context. The observation schemes adopted in this study were the Communicative Orientation of Language Teaching (COLT) (Spada & Fröhlich, 1995), University of Cambridge Classroom Observation Schedule (UCOS) to focus on what teachers and students actually do in the classroom and how they interact (Allen, Fröhlich, & Spada, 1984: 232), and a self-made checklist.

The researcher observed 10 EFL teachers (teaching English First Paper) both in rural and urban sites. The main aim of this section is to present and discuss the findings of the classroom observation using the COLT, UCOS, and a self-made checklist. Although these instruments focus on describing the instructional practices, procedures and materials in the foreign language classrooms, COLT had a more general application while UCOS had been designed to capture features salient to examination preparation courses. This section also discusses the analysis of additional categories defined for this study, as the purpose of classroom observation was also to obtain a view of the climate and rapport together with the interaction and functioning of the classes. In the first part of this section, information from COLT (part-A) is provided, covering interaction, control of the content of the lessons, potential predominance of teacher fronted activities, most common skills used by the students and materials employed. The analysis using the UCOS (part-2) provides information on occurrence of activities which might be expected in examination classes, the types of texts actually used in the classes, class time spent on grammar and vocabulary activities and classification of reading, writing, listening and speaking activities.

This is followed by a further analysis of the observation through an additional self –made checklist which covers a number of things that the teachers talked about the HSC examination in English, strategies recorded throughout the lessons, teacher-student interaction not covered by COLT and UCOS, sources of the materials used on the preparation courses and the extent to which the teacher

adapted the materials to suit the specific needs of the class, topics appearing in the materials used, homework and instances of laughter or shouting as an indication of the overall atmosphere. This collection and detailed analysis of the activities of the classrooms was used for two purposes - to gather information about the nature of the HSC examination preparation classes, and to provide data to inform of the discussion of the washback effect of the test on each of the ten classes observed.

In this section of the study, the evidence of washback was sought in various ways: (a) the nature and focus of the classroom activities and instruction, (b) the type and content of instruction, (c) the amount of language instruction, (d) the amount of exam-related instruction, and (e) the type and origin of the classroom materials and the atmosphere of the class.

5.2.2 Profile of the Participants

All the 10 observation participants were currently EFL teachers working at ten different colleges both in urban and rural areas. Of the 10 participants, 4 were females and 6 were males. Each has a teaching experience of more than 10 years (Table 5.65). All of them received a master's degree in English. Their teaching hours ranged from 8-12 hours per week. At the time of observation, they were teaching HSC students with similar level of proficiency in English. None of them had experience of studying or working abroad. Two participants reported having received teacher training in ELT, and one teacher claimed to have been exposed to task-based activities:

Table 5.65: General characteristics of the participants observed

General Characteristics of the Participants	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
Sex	M	F	M	F	F	M	M	F	M	M
No. of years of teaching experience	15	19	13	10	14	15	12	17	11	9
No. of teaching hours per week	10	12	8	12	10	12	8	10	12	12
Class size (No. of students in class)	49	55	74	62	50	42	56	63	51	77
Experience of being in an English-speaking country	No	No	No	No	No	No	No	No	No	No
Training in teaching methodology	Yes	No	Yes	No	No	No	No	No	No	No

5.2.3 Classroom Observation Schedule- COLT (Part-A)

The COLT consisted of two parts. Part-A of the COLT was employed in this study, as a classroom analysis at the level of activity matched the nature of the research questions to be answered. Part-B was not employed since the focus of this study was not mainly on the language used in the class.

The classroom observation examined the washback effect on the teaching and learning. The categories based on Part -A of the COLT were designed to (a) capture significant features of classroom events, and (b) provide a means of classroom interaction. The main focuses in this phase of the study are related to the washback effect on:

   a) the English language syllabus at the HSC level,
   b) textbook materials used in practice,
   c) teachers' teaching behaviours, and
   d) teachers' beliefs, attitudes and perception related to test.

The classroom activities were designed to describe in order to investigate such aspects as whether the lesson was student-centred or teacher-centred, how many learning opportunities were provided, and what pedagogical materials teachers used in teaching, e.g. real-life materials, main textbook (English for Today) or practice examination papers.

The observation scheme (COLT) for the study consisted of five major categories including time, participant organisation, activity type, content, and material used. They were all coded in the classes. The present researcher ticked under the category of participant organisation and materials used during the observation, but made noted under the category of time, activity type and activity content. The major categories are briefly discussed below:

   Time: How is time segmented within the lesson as a percentage of class time? This category related to instructional behaviours in the classroom. The unit of analysis chosen was a 'segment'. A segment is defined by Mitchell (1988) as "a stretch of classroom discourse having a particular topic and involving participants (both the teacher and students) in carrying out an activity or task through interaction" (p.12-14).   A change of topic/activity type or a mode of interaction

indicates a completion or the start of a new segment (Gibbons, 2006, p, 95). The segment was selected as the basic unit of analysis because it has distinctive features, both linguistic and pedagogic, and therefore can be readily divided into categories as a percentage of class time. Segment boundaries were identified on the basis of 'focusing moves' and 'framing moves' (Gibbons, 2006), which were indicators of the completion of one stage of a lesson and the beginning of another. Therefore, the first step in analysing any lessons observed was to divide the lesson into segments.

Participant Organisation: Who is holding the floor/talking during the segments of the lesson as a percentage of class time? Participant organisation covers three basic patterns of organisation for classroom interactions. The three patterns are: whole class – involving teacher to students, or student to students, pair work or group work, and individual work (Allen et al., 1984).

These categories describe how a lesson is carried out in terms of the participants in the classroom interaction. The categories reflect different theoretical approaches to teaching. Moreover, student talk in a teacher-centred classroom is frequently limited to the production of isolated sentences, which are assessed for their grammatical accuracy rather than for their communicative competence. Highly controlled, teacher-centred approaches are thought to impose restrictions on the growth of students' productive activity. Participant organisation is one of the rationales behind the imposition of language tests in order to encourage more practice opportunities for students. Therefore, it is necessary to observe the participant organisation of classroom interaction patterns in this study. The findings enabled a comparative investigation of the interaction patterns in classes to see if there were any differences between different groups.

Activity Type: What are teaching and learning activities realised through various tasks and activities as a percentage of class time? After each lesson had been segmented and interaction patterns of classroom activity analysed, the aim was to look more closely at the types of activity carried out within the segments. Each activity was separately noted down such as discussing, lecturing, or singing.

Content: What are the teacher and the students talking, reading and writing about? Or what are they listening to? 'Content' refers to the subject matter of the activities.

Materials Used: What types of teaching materials are used and for what purpose?

Types of materials:

a) The study looked at the written materials, such as textbooks, worksheets, and mock examination papers;

b) It examined whether any audio materials, such as songs were used in the class; and

c) It observed if visual materials, such as films were used.

Purposes of materials:

a) The study examined the pedagogical (e.g., main textbook specifically designed for EFL learning) purposes of using the materials;

b) It investigated the semi-pedagogical (e.g. model examination papers) purposes; and

c) It checked the non-pedagogical (materials originally intended for non-teaching purposes, such as English songs and films) purposes of using the materials.

As mentioned, the lessons of each class were coded according to COLT (Part-A). The basic units of analysis for this part of the observation scheme are 'activities' and/or 'episodes'. Activities and episodes are the units which form the instructional segments of the lesson. Activities consist of one or more episodes and mark changes in the category of the features of COLT being observed.

5.2.3.1 Participant Organisation

Three basic patterns were observed whether the teacher was working with the whole class or not, whether the students were divided into groups or they were engaged in individual work, and whether they were engaged in-group work and how it was organised. The findings are represented in the Table 5.66. The first COLT category looked at whether classroom activity focused on the teacher or on the students working as a whole class, in groups or as individuals. These categories described how a lesson was carried out in terms of the participants in the classroom interaction. The categories reflected different theoretical approaches to teaching. For example, Allen et al. (1984) and Gibbons (2006) consider group work as an important factor in the development of fluency skills. Allen et al. (1984, p. 236)

claim that "In the classes dominated by the teacher, students spend most of their time responding to questions and rarely initiate speech". Moreover, student talk in a teacher-centred classroom is frequently limited to the production of isolated sentences, which are assessed for their grammatical accuracy rather than for their communicative competence. The details of the participation organization are presented in Table 5.66. For the purposes of this study, the ten teachers were anonymous and coded as T1 to T10. The class duration was 50 minutes. The participant organisation patterns maintained in the study were (a) teacher to students (pre-lesson activities, lecturing, describing, explaining, narrating, directing, checking answers for exercises together, practising test, reading aloud), (b) individual work (student-student), (c) group work(students are working on a certain task in groups), and (d) pair work (sharing one another, e.g. on dialogue, problem solving, etc). The findings are presented in the above table (Table 5.66):

Table 5.66: Distribution of (%) participant organization
[M=Mean, STVD= Standard Deviation]

Participation Organization	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	M	STDV
Teacher to students (class)	71	67	51	44	78	58	79	77	74	76	67.5	12.36
Individual work	9	14	23	17	11	19	11	12	12	10	13.8	4.45
Group work	7	11	6	15	3	16	5	8	9	9	8.9	4.0
Pair work	13	8	20	24	8	7	5	3	5	5	9.3	7.0
Total (%)	100	100	100	100	100	100	100	100	100	100	Average	

The classroom observation found that teachers used maximum time of the class. It indicates that the teacher was the main focus of the lessons. It further proved that the class was teacher-centered. On average, more than half (67.5%) of the total class time was used by the teachers, while another 13.8% of the time involved individual work and tasks (including exchange of views). The other interaction included a number of practice tests, resulting naturally in individual students working on a single task.

It was found that T7 used 79% of class time, the highest amount, for his classroom teaching, whereas T4 used only 44% of class time, the lowest span of time. She (T4) used a considerable amount (24%) for pair work involving her

students in a number of activities. With regard to participant organisation, the study found that most of the teachers (90%) occupied maximum class time indicating that the classroom was teacher-dominated rather than student-oriented. This practice is directly opposed to communicative language teaching (CLT). However, it was appreciative that T4's class was student- oriented one. She used the target language in the class, and involved the students in the classroom activities. Activity types were grouped into (1) teacher activities, (2) teacher and student activities carried out together and (3) student activities. Each activity was classified, such as discussion, drill or singing. The averaged participation (percentage) as well as the individual teacher's class-time occupation is also shown in the Figure 5.83:

Figure 5.83: Teachers' class participation organization
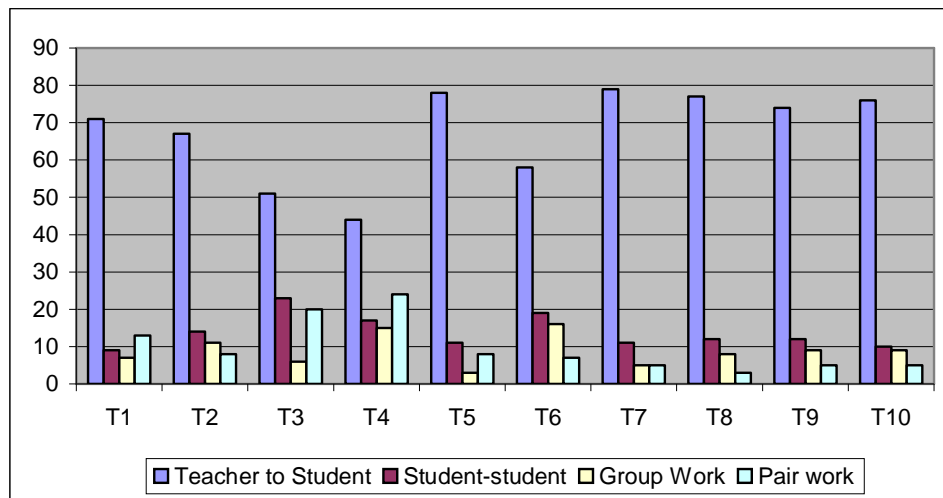
The present EFL curriculum has introduced CLT, and the textbook (*English for Today*) materials have been designed and developed in such a manner that, it can ensure practice in four basic skills of English language: listening, speaking, reading, and writing. Classes are expected to be interactive with students actively participating in the classroom activities through pair work, group work, and individual work. But in reality, EFL teachers failed to achieve desired objectives set by the syllabus and curriculum. The figure (Figure 5.84) shows that an average of 67.5% class time was used by teachers, nearly 14% of the time was spent in 'student to student' interaction (e.g. dialogues, conversation, asking question, personal talk), approximately 9% class time was used for groups, and more than 9% time was utilized for pair work:

Figure 5.84: Average class participant organizations
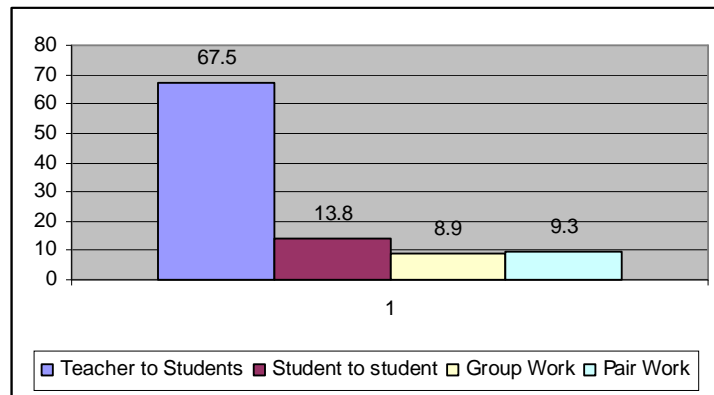
5.2.3.2 Classroom Activity and Content

The purpose of looking at activity type in classroom teaching was to explore what kinds of teaching and learning were realised through various activities. By investigating the content of the activities carried out in the classroom, the researcher explored the subject matter of the activities - what the teachers and the students were talking, reading, or writing about, or what they were listening to. Activity types were grouped into teacher activities and student activities. Findings relating to the content were again reported as a percentage of class time. The analysis of the ten classes of the 10 teachers (Table 5.67) showed (a) what types of activity were carried out in the lessons and how lessons were segmented according to the percentage of time devoted to them by the four teachers, and (b) who was holding the floor and in what ways.

COLT identifies the content of the classroom activities, measuring where the focus lies – on meaning, form or a combination. The two main categories are topics related to classroom management (procedure) and language issues. There is also a category which provides a binary distinction to be made about whether the content refers to the immediate classroom and the students' immediate environment (*Narrow)*, (the discussion focussing on Narrow subjects was limited to a brief discussion about their feelings about the results of a test and describing their important friendships, etc.) or encompasses *broader topics* (Broad). Analysis of participant organisation indicated the predominance of teacher-fronted activities. This is reflected in content in the subcategory *Procedure,* which took up on average 12.7% of the class time.

285

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

The largest content area was the sub-category broad, (i.e. the discussion of topics outside the immediate concern of the classroom, HSC examination related) and a significant amount of the class time categorised in this way was a reflection of the time the teacher spent speaking about the examination. The categories of procedure and broad accounted for nearly (12.7+64.1) 77% of the total class time. Only slightly over 19% of the class time was spent on aspects of language teaching and learning (vocabulary, pronunciation, grammar, discourse, function, sociolinguistics, etc). Information about written discourse was the most significant language focus, followed by vocabulary, and the combination of discourse and vocabulary, which was typically work, related to discourse markers (Table 5.67).

The discussion focusing on narrow subjects was limited to a brief discussion about their feelings about the results of a test. Language instruction played a significant role in the observed classes. Activities focusing on both vocabulary and grammar were the most common category of classroom content. The learning of vocabulary was particularly important. The teacher and students spent some of the time working on new words, collocations and phrases. The broad items included the discussion of topics outside the immediate concern of the class room, test, materials, seriousness, counseling, etc. The present study found that more than 61% was spent for the broad items. It was also found that T10 used 67.26% as the highest amount of time spent for Broad topic, whereas T4 used 50% class time, the lowest amount of time for *Broad* purpose. The table below (Table 5.67) presents the details of classroom activities and contents taught:

Table 5.67: Content of lessons as a percentage of total class time

Content	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	M	STDV
Procedural Directives	17.54	8.85	9.9	8	10.33	12.5	11.5	13.25	19.25	16.5	12.7	3.84
Vocabulary	1.95	14.08	5.22	7	9.32	8.2	7.5	5.5	2.35	2.5	6.3	3.75
Pronunciation	0.1	1.53	1.48	2	0.89	1.1	2	1.25	0	1.25	0.88	0.68
Grammar	1.17	1.64	4.79	3.3	3.11	4.5	3.75	4.5	3.25	4.24	3.1	1.2
Spelling	0	0	0.33	0.5	1.25	1.0	1	1.5	0.25	0	0.6	0.55
Function	1.48	1.05	1.27	7	1.5	1.5	1.5	1.75	0	1.25	1.4	1.88
Discourse	4.09	0.93	4.61	3.5	2.5	2.1	3.25	0	2.75	0	2.6	1.6
Sociolinguistic	0	0.05	0.22	10	0	0.5	0	0	1	0.25	0.3	3.1
Vocabulary and Discourse	1.62	2.62.	0.64	1	1.88	2.2	2	4.25	1.75	2	1.9	0.98
Vocabulary and Grammar	0.27	15.2	0.88	3.3	1.43	2.5	2.25	1.75	2	0.5	2.3	4.39
Narrow	4.6	0.14	2.22	4.4	4.6	3.5	4.25	3.5	3.25	4.25	3.7	1.39
Broad	67.18	53.91	68.44	50	63.19	60.4	61	62.75	64.15	67.26	64.1	5.9
Content total	100	100	100	100	100	100	100	100	100	100	**Average**	

In the teacher dominated language classroom, a little learning takes place. During the observation, the present researcher found that the teachers were playing dominating role for the examination preparation activities. It was found that the teachers spent most of the class time for Broad, and Procedural purposes. The figure below (Figure 5.85) displays the findings of the classroom activities and contents:

Figure 5.85: Projection of lesson contents
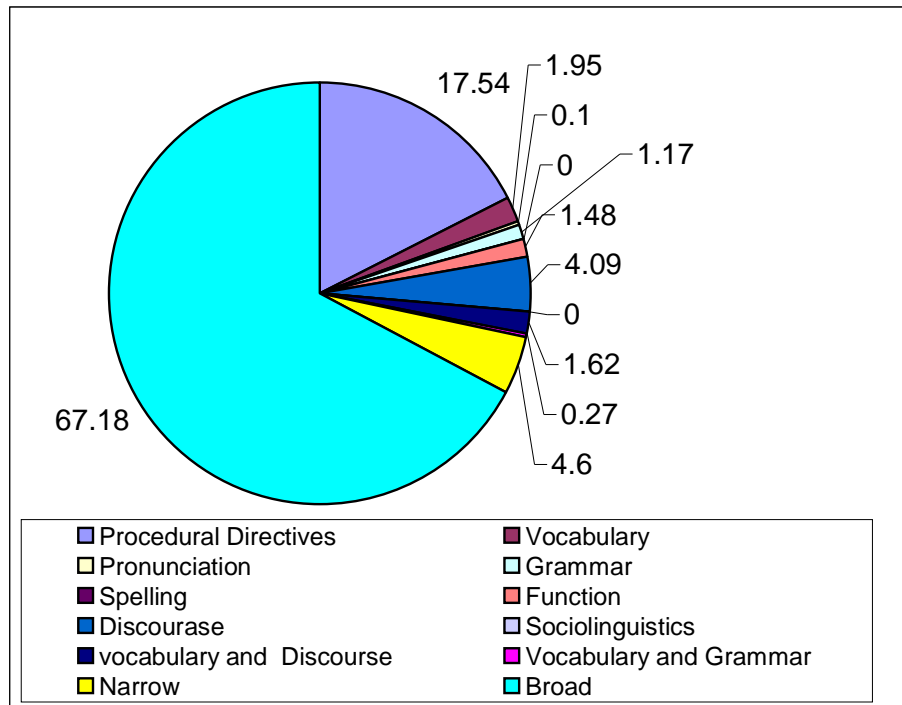
Broad topics occupied the major part of the class which was mainly the test; and this was not the really concern of the class. Vocabulary and grammar references were more prominent in Writing. The main focus in all 10 classes was on meaning with emphasis on discussion of broad topics. There was little focus on Narrow topics (almost absent in T2), which was to be expected, considering that the classes were meant for students and the focus of the course was HSC English syllabus, a topic which itself was classified as Broad as although it was the focus of the class, the test was an event outside the classroom.

The teaching of language played a less significant role in all observed classes. A considerable part of the lessons in T2 was spent focusing on language, in particular vocabulary and vocabulary (16%) in combination with grammar. However, the teaching of vocabulary, pronunciation and grammar in the classes of

287

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

all teachers took up considerably more time compared to other task. T4 was found more active than other teachers. She used 20% of class time for teaching direct communication (function, discourse, and sociolinguistics) purposes.

5.2.3.2.1 Content Control of Classroom Activities

In order to assess the level of involvement of the students in the control of the lesson, the researcher (using COLT) identified who was responsible for content selection. The variables in this category were the teacher, the student/s, the teacher and text, or a combination. The teacher's individual control over the class as well as the average percentage is reflected in the following table (Table 5.68). In the class, it was found that average more that 75% control lied with the teacher and their choice of the text. For an additional 25% (approximate) of the class time the students shared control of the content of the lessons with the teacher, for example when the teacher asked the students to share their experience of sitting for the HSC examination, or how difficult they found a particular exercise. At no time did the students alone decide on the content of the classes.

Reflecting the amount of student involvement in all 10 classes is shown in the following table (Table 5.68) and figure (Figure 5.86). The control of the content of the classroom activity was most commonly shared between the teacher, text and students and varied from 55% to 90%. For example, the teacher presented a text, and explained the exercise, and then allowed the students to work in pairs or small groups to work through it together. The student control varied from 10% to 45% in different classes. The highest covered teacher- controlled classroom activities were found (90%) in the class of T7, whereas T4's class was the lowest teacher controlled (52%). T7 was mostly occupied with the text and himself. He explained the text, tasks, and exercises on his own ways and sometimes (10%) asked his students whether they understood. Table 5.68 presents the average and individual results of content control expressed as percentage of each total class time for the three classes:

Table 5.68: Content control as a percentage of total class time

Content Control	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	Mean	STDV
Teacher/text	67	80	71	52	79	77	90	82	73	78	74.9	10.22
Teacher/text/student	33	20	29	48	21	23	10	18	27	22	25.1	10.2247
Total	100	100	100	100	100	100	100	100	100	100	Average	

Communicative language teaching (CLT) requires students' direct and active involvement for developing communicative competence, but the present researcher found it absent from the classroom. The study found that almost 90% teachers tried to control the contents, tasks, and activities for the cause of examination preparation. It is believed that the influence of examination leads the teachers to control the contents and classroom activities. It was found that the teachers talked about the HSC examination, and taught them how to prepare their students for the test. Teachers' content-control was found high because the negative washback influenced their personal and academic behaviours. The figure below shows how the content control occurs in the language classroom:

Figure 5.86: Content control as a percentage of total class time
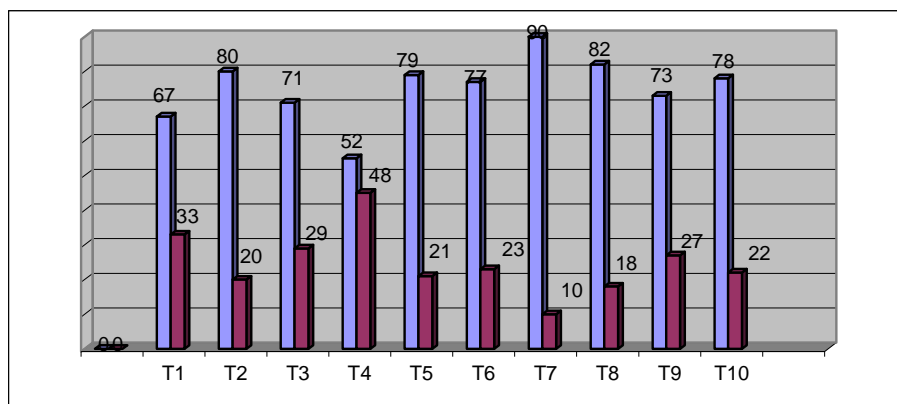
5.2.3.2.2 Student Modality

Identifying the skills the students were involved in during the classroom activities is recorded in the section called 'Student Modality'. This is broken down into the four skills (e.g. listening, speaking, reading and writing) with a fifth category which allows activities such as drawing or acting to be recorded.

Writing was the most common skill used by the students in the classes of all 10 observed teachers, representing average 51.5% of total class time. In some cases, while practising HSC practice/model tests, they were mostly listening to the teacher explaining procedure, giving information related to HSC examination or checking answers to practice test materials. Again writing in combination with listening was the second most common modality at an average of 7.6% of the total class time. Details of the student modality are shown in the in the table (Table 5.69) below:

Table 5.69: Student modality as a percentage of total class time
(M=Mean, STDV=Standard Deviation)

Student Modality	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	M	STDV
Writing only %	48	52	42	55	56	59	52	54	48	49	51.5	4.9
Speaking only %	4	3	10	8	5	8	10	10	3	17	7.3	4.7
Reading only %	4	2	5	4	7	4	5	5	3	5	4.4	1.3
Listening only %	7	7	7	2	6	9	6	5	11	5	6.5	2.4
L + S %	4	6	7	6	5	5	5	5	10	0	5.3	2.5
L + R %	4	12	9	2	5	5	6	2	6	8	5.9	3.1
L + W %	15	5	10	2	6	2	10	10	8	3	7.6	4.9
S + R %	2	3	2	5	3	3	3	3	4	5	3.3	1.0
L + S + W %	10	5	4	8	4	2	2	3	3	8	4.9	2.8
L + S + R %	2	5	4	8	3	3	1	3	4	0	3.3	2.2
Total (%)	100	100	100	100	100	100	100	100	100	100	100	

Speaking was the third common modality at an average of 7.3% of the total class time. Furthermore, speaking and listening jointly took at an average of 5.3% of total class time. Listening plus speaking plus writing (4.9%) indicated activities where students exchanged information and took notes, and speaking plus reading (3.3%) was used when students were reading and summarising information to a partner:

Figure 5.87: Students' involvement in language practice
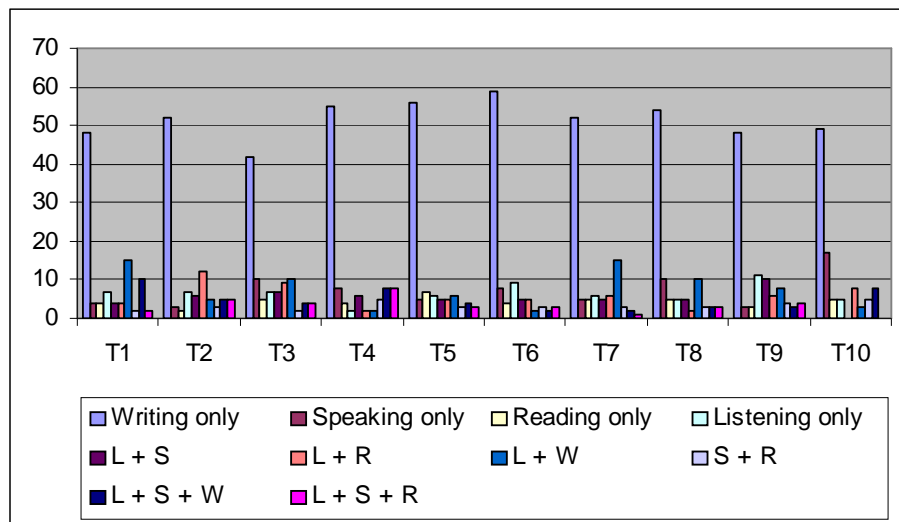
Writing was the most common skill in the all classes, representing on average (51.5 + 7.6+ 4.9) 64% of the total class time. Figure 5.87 displays the averaged results of student modality expressed as percentage of each total class time for the 10 classes. Writing, both alone and in combination with other skills, was the most common skill used by students at all schools. In general, students of T4 used a

290

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

broader range of skills and covered the four skills more evenly. The classroom observation found that students took part in writing to large extent. The teachers made them practise writing as an individual activity as well as a combined activity with other tasks.   Writing gets the highest priority in the classroom because it is mainly tested in the HSC examination. The findings of student modality are supported by the questionnaire surveys which found that writing and some other linguistic elements were taught because they (skills and elements) were tested in the examination. The classroom activities and academic behaviours of the teachers and the students were guided by the influence of the HSC examination. The findings adequately proved that washback of the HSC examination influenced classroom teaching and learning.

5.2.3.3 Materials Used in the EFL Class

This section presents and discusses the findings related to the materials used in the classroom teaching. The present researcher used COLT and recorded significant features about the materials used during the class. The type of text was broken down into length with short pieces of written texts, for example single sentences or captions, being labelled as 'minimal' and longer ones 'extended'. The origin of the material was also considered important. The researcher carefully observed and identified the materials being used in the language classroom.

The classroom observation checked if any authentic materials were used. Whether any adaptations made to materials were also noted in this section. It was found that more than 80% teachers were heavily dependent on the commercially produced written materials such as guide book, suggestion book, test papers, etc. The 30% observed teachers did not use *English for Today* (for classes 11-12) written by the NCTB at all. The types of teaching materials for all teachers were not substantially different. There was an impact of the HSC examination in English on teaching materials. It was found that most of the teachers used test-oriented commercially produced materials.

The researcher found that three teachers (T1, T3 and T5) used 75% class time practicing examination related materials. T1 and T3 used model tests book, and T5 used suggestion book. Only T4, unlike other participants who attached

more importance to language forms, stressed the development of students' ability to use English. She (T4) was so highly motivated that she spontaneously experimented with communicative activities as well as cooperative learning activities (e.g., pair work/ group work, language games, questions and answers) in her classes. Not only was she observed frequently utilizing authentic materials, but she was also found using *English for Today* (for classes 11-12) more creatively and trying hard to encourage her students to interact in class. Table 5.70 presents the categories of materials used in the class. The materials are abbreviated for the benefit of presentation in the table (Table 5.70).

[**Key**: EFT=*English for Today* (for classes 11-12) ● GB= Guide Book ● TP=Test Papers● PQ=Past Questions ● AM= Authentic Materials (Newspaper article, Cultural current events, etc) ● RM=Reference Materials]

[Symbols √ = ± 25% class time, √√= 50% class time, √√√ 75%= class time, √√√√=100% class time]

Table 5.70: Teacher's use of materials as a percentage of total class time

Teachers	EFT	PQ	TP	GB	AM	Audio	Visual	RM
T1	√	√	√√√ (75%)			-	-	
T2		√√ (50%)	√	√		-	-	√
T3		√√ (50%)	√√√ (75%)			-	-	
T4	√√ (50%)				√	√	√	√
T5	√			√√√ (75%)		-	-	√
T6	√		√√ 50%	√		-	-	
T7			√	√√(50%)	√	-	-	
T8	√√ (50%)		√√ (50%)			-	-	
T9	√√√ √(100%)					-	-	
T10	√√ 50%	√		√		-	-	√

It was found that only T9 used *English for Today* during in the whole class, whereas T4, T8 and T10 used *English for Today* half of the class time. The four teachers (T1, T3, T6 and T8) spent considerable amount of time teaching commercially produced test papers. T4 used *English for Today*, authentic materials, some audio-visuals. She also mentioned some reference books (e.g. Oxford Dictionary) in the class. It was found that T2. T3 and T7 never used *English for Today* during the whole class period.

The study found that test papers and past questions were the most common type of materials used in almost all classes (90%). Some combinations of material

types were only found in some of the classes. The observed EFL teachers used commercially produced test-related materials for the preparation of the HSC examination in English. The findings are supported by the results of the questionnaire surveys which found that most of the teachers used commercially produced materials and avoided *English for Today*. The study is further supported by Cheng (2004) who found in China that 80% teachers and learners used commercially produced materials for the preparation of College English Test (CET).

Though language use was not specifically recorded; the observer noted that students were more likely to use their own language (Bengali) during class. There are several possible reasons to account for it. For instance, there were larger numbers of students in every classes; it meant that there was potentially more opportunity for students to congregate and share ideas in their native/ common language. Large classes are also more difficult to monitor than smaller groups if the teachers had decided that they preferred the use of English in the class. The mixture of students from different background in those observed classes may also have been responsible for the students using their first language as they struggled to follow the class. The findings of materials used in the class derived from COLT observed the evidence of negative washback on teaching and learning in general and on the use of materials in particular.

5.2.4 Classroom Observation Schedule- UCOS

The present researcher used a modified version of UCOS for the present study. The UCOS 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 texts used in each of the classes were also recorded (using COLT). A large part of the UCOS focused on what skills the students were using in the classroom. Here, UCOS gives much more detail that the 'modality' category of COLT by describing the activity. The original UCOS was adapted to the purposes of this study, as the existing categories did not always comprehensively reflect what happened in the classrooms (Appendix- 2B). The Modified UCOS contained a broad list of possible task and text types. However, it was found that a large number of the texts actually used in the classes did not fit into the existing categories and were therefore recorded as additional categories.

Initially, anything that occurred in the classrooms that did not fit under the existing classifications was listed separately (self- made observation checklists). Similar activities were used to form a new category which was added to the instrument under the existing framework. In other instances, categories mentioned in the UCOS were not observed, and these were eventually deleted from the instrument. This category focused on the teachers and recorded activities which might be expected in HSC examination preparation classes. Overall examination-related activities of the total class time are shown in the table (Table 5.71). On an average, the teachers gave the students direct practice of HSC examination for 17.5 minutes in a 50-minute class. Examination-related activities altogether occupied for almost 42 minutes. The teachers most commonly gave the students feedback on reading and writing tests by giving the answers and explaining where in the text they could be found.

The students were sometimes encouraged to reflect on their performance on the practice tests and to initiate the necessary additional study. The individual teacher's examination-related activities were accounted separately. The students also spent some of the total class time completing tasks under examination condition. Reviewing answers to reading comprehension or writing tasks was a common activity. The findings of the examination activities are presented in the table (Table 5.71) below:

Table 5.71: Examination-related activities of total class time

Exam Related Activities (ERA)		Total time spent for each activity (minutes)										Average
		T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	Minutes
ERA- 1	Teacher gives the students tasks under exam conditions	21	18	19	12	15	16	21	11	23	15	17.5
ERA-2	Teacher gives the students the test to do at home (self-timed)	3	10	11	9	11	15	12	12	10	10	10.3
ERA-3	Teacher gives students feedback in the form of HSC	2	1.5	2.7	2.5	2	2	1.5	2	1.7	2	2
ERA-4	Teacher gives feedback on student performance item by item (T gives the right answer without explanation of reasons)	2	2.5	3	1	1.5	2.5	2.7	3	1	2	2.12
ERA-5	Teacher identifies	10	12	10	0	0	4	5	5	6	7	5.9

	answers in a text (Reading or Listening) and explains										
ERA-6	Teacher asks students to consider their strengths and weaknesses with respect to the test requirements	5	0	0	4	4	3	0	0	3	5	2.4
ERA-7	Teacher sets tasks under strict time pressure	0	0	2	3	2	2	2	2.5	2.3	3	1.88
Total % of examination-related activities		43	44	47.7	31.5	35.5	44.5	44.25	35.5	47	44	41.7

The classroom observation found the teachers providing the answers, identifying the answers in the text. The teacher at times supplied answers after the students had spent some time discussing the task in the whole class (some time in groups or pair) and reaching some form of agreement. T9 used the highest amount of the examination-related activities which was 23 minutes as a single activity; altogether he spent 47 minutes out of a total 50-minute class. The figure (Figure 5.88) below reflects the findings:

Figure 5.88: Examination related activities
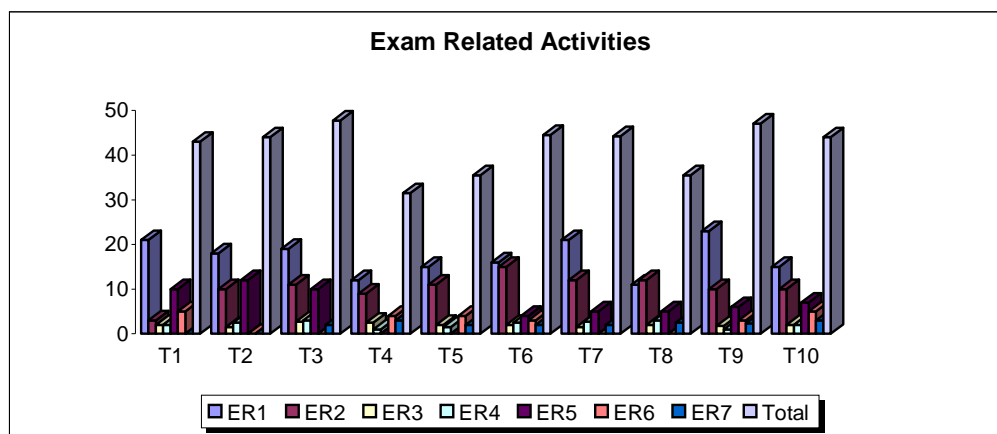
5.2.5 The Self-made Checklist (Further Analysis)

An observation checklist was applied to recoding some activities during the lessons which were not specifically identified by either COLT or UCOS. The findings from the checklist are now presented. Through the self-made checklist, teachers' personality and professional behaviours were coded. Teacher's personality and professional behaviours contribute learning or not learning. Learners'

concentration and classroom performance largely depends on the teachers' personality, attitudes, and on the amicable relationship (Turner, 2008). Wang (2010) found strong influence of teachers' factors in contributing to generate positive or negative washback in varying degrees. The researcher observed 10 EFL teachers. The findings of additional analysis are present in table (Table 5.72) below:

[Teacher's personality and professional behaviours are coded as, A= always, E=Excellent, F=Frequently, G=Good, M=Moderate, N=No, P=Poor, S= Sometimes, Y=Yes]

Table 5.72: Teachers' personality and professional factors in generating washback

Personality & Professionalism	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
Friendly (Y/N/M)	M	Y	N	Y	M	Y	N	Y	M	Y
Angry (Y/N/M)	N	M	Y	N	N	M	Y	N	Y	N
Introvert (Y/N/M)	Y	M	Y	N	N	N	M	N	Y	M
Extrovert (Y/N/M)	N	M	M	Y	Y	Y	M	Y	N	M
Laughter (Y/N/M)	Y	Y	N	Y	Y	Y	N	Y	N	M
Shouting (Y/N/M)	N	N	Y	N	N	N	Y	N	Y	Y
Smiling (Y/N/M)	Y	Y	M	Y	Y	Y	N	Y	M	M
Well-behaved (Y/N/M)	Y	Y	M	Y	Y	Y	M	M	M	M
Encouraging (Y/N/M)	M	M	M	Y	M	Y	M	M	N	M
Sincere & Caring (Y/N/M)	M	M	M	Y	Y	M	M	Y	N	M
Punctual (Y/N/M)	M	M	Y	Y	M	Y	N	N	Y	M
Fluent (Y/N/M) in English	M	M	M	Y	Y	Y	M	M	N	N
Knowledge of Communicative Competence (E/G/M/P)	P	M	P	G	G	G	M	P	M	P
Target language use (A/F/S/N)	S	S	S	**A**	F	A	S	S	S	S
Presentation(E/G/M/P)	G	M	M	E	P	G	M	M	M	G
Pronunciation (E/G/M/P)	G	G	M	G	M	G	P	M	M	P
Preparedness for Teaching (E/G/M/P)	G	M	M	E	M	G	P	P	M	P
Hesitant (Y/N/M)	M	N	Y	N	N	N	Y	Y	Y	N
Curriculum Knowledge (E/G/M/P)	M	P	M	E	G	E	P	P	P	P

As the Table 5.72 shows, 6 teachers (T2, T4, T6, T8 and T10) were found friendly to their students. Andrew (2004) suggests that friendly teachers are always considered as good language teachers. It is sometime true that all successful language teachers are not treated as socially amiable. Out of 10 teachers, 3 teachers (73, T7 and T9) were found very angry in different situation while teaching their students. Some teachers (T1, T3, and T9) were found very introverted while

teaching. The introverted teachers taught to the test and less friendly to their students. Four teachers (T4, T5, T6 and T8) were found very extrovert. The extroverted teachers were found friendly. Ellis, R (2001) finds extroverts as good instructors. This study found that the extroverts were better teachers than others. Three teachers (T4, T5 and T8) were found very sincere and caring to their teaching. Among the 10 observed teachers, two (T4, T5 and T6) were found fluent in English at satisfactory level. Only three teachers (T4, T5 and T6) had good knowledge and experience of communicative competence.

The observation recorded that only three teachers had good level of curriculum knowledge. One of the interviewed teachers commented that curriculum knowledge was not important to teach English to his students. Teachers' perception on largely contributes to generation of negative washback on teaching and learning (Andrews, S., & Fullilove, J. 1994). The study found that only 34 teachers (T4, T5, and T6) were found well informed of the goals and objectives of the syllabus and curriculum while others had very poor or moderate level of curriculum knowledge. Chen (2002), in her study, finds that teachers prefer to 'teach to the test' when they have little knowledge of curriculum goals; and therefore, they use commercially produced materials for test preparation. Promotion of beneficial washback has deep relation with teaching to the curriculum opposed to teaching to the test.

Noble and Smith (1994) point out those teachers' manners and professional behaviours are the indicators of being good language teachers. The observation schedules (COLT, and UCOS) and a self-made checklist were complementary for each other. The classroom observation found sufficient evidence of negative washback of the HSC examination in English on teaching and learning English as a foreign language.

5.2.6 Summary of the Results of Classroom Observation

The use of COLT and UCOS in combination with the specific further analysis (self-made checklist) enabled the present researcher to collect qualitative primary data from the respondents for the present study. This was an attempt not only to determine the range of activities that might occur in the HSC examination preparation class, but also to identify the amount of lesson time in which the students in the observed classes were actively communicating, as this would be an

indication of good classroom practice which could in turn possibly be seen as a result of a 'good' test. Teachers' and students' perspectives were elicited and cross-referenced to the findings of the instruments, using a combination of purpose built questionnaires and interviews. The combination of the instruments used to draw a possible true picture of influence of the HSC examination in English and EFL education.

All 10 observed classes were found to consist predominantly of materials written for language students; contained a significant number of practice tests; included examination-related activities; and incorporated few academic study skills. Two of the books mentioned in the materials analysis section (Section 5.3) were found to be examples of a more traditional approach to test preparation, which focused on familiarising students with the test and providing opportunities for test practice both in and out of the class. Normally, T4 incorporated a communicative methodology, included elements of language development and gave the students practice with a number of academic study skills. Most of the teachers (80%) were totally HSC examination focused, i.e., not preparing students for academic study. It should be noted, however, that with different teaching backgrounds, beliefs and personal teaching styles notwithstanding, each of the teachers had a certain amount of material that they were required or expected to get through in the limited time-frame of the course.

The data presented above give an idea of the participants' beliefs and some scenarios of their teaching practices in the classroom. On the whole, all the participants were interactive and cooperative. They all impressed the researcher as committed and responsible EFL teachers although their conceptions of teaching, their levels of language proficiency (e.g., competence in terms of four skills, awareness of the socio-cultural aspects of language and language use as well as knowledge of pedagogy), the ways in which they conducted their lessons, and their devotion to work differ to varying degrees.

The findings revealed that due to college differences as well as differences among teachers and students, not only the ways teachers perceived and reacted to the HSC examination and its washback varied from college to college, but they also differ from individual to individual. On the one hand, teachers' beliefs and knowledge of the HSC examination vary from context to context. When talking

about the effects of the HSC examination in English on their teaching, the majority of them suggested that that they were motivated by the test. They also expressed in a way that the examination preparation was their prime concern. Out of 10 observed teachers, a number of 7 teachers could not make any difference between teaching to the test and teaching to the syllabus, which could be interpreted that their curriculum knowledge was indeed limited or insufficient. One significant feature that emerged from the data was that the observed teachers seemed to be more nervous about the HSC examination in English subject.

The overall findings of the classroom observation reflected that the HSC examination in English influenced most of the teachers directly. But T4 was found an exception in this case. The EFL classes were found teacher-centered and teacher dominated. On an average, 67.5% of the total class time was occupied by the teachers. They dominated class time, contents, and class activities through different types of actions. The classroom observation revealed that some of the teachers used mainly the grammar-translation method. For instance, one teacher, in her class, asked her students to translate sentences from Bengali into English to ensure that the students fully mastered the structure and its meaning. To a certain degree, the use of the grammar-translation method was counterproductive; not promoting students' communicative skills, especially speaking skill, as prescribed in the syllabus.

Mostly, writing and reading comprehension were practised in the class because it was considered to be the demand of the test. An average of more than 64% class content-control was exclusively with the teachers under broad category. As a single activity, 51.5% class time was spent for teaching and practicing the writing skills. Another considerable amount of time was spent for writing along with the combination of other skills e.g. writing while listening (51.5 + 7.6+ 4.9). Therefore, the writing skills claimed 64% of the total class time. There were little opportunity of practicing speaking and listening. There were very little opportunities for pair work and group work in the observed classes, except in the class of T4. With regard to use of materials, the classroom observation found that more than 80% teachers were reliant on the test related materials though a few teachers occasionally used *English for Today* (EFT) in the class.

It was also found that (using UCOS) nearly, on an average, 42 minutes (out of 50- minute class) was spent for examination preparation activities (EPA). The

teacher at times supplied answers after the students had spent some time discussing the task in groups and reaching some form of agreement. T9 taught the highest amount of the Examination-related activities which was 23 minutes as a single activity; altogether he spent 47 minutes out of a total 50-minute period.

With regard to the instances and ways of mentioning the HSC examination, all teachers referred to the HSC examination frequently during the class. They advised their students in many ways to be more serious about the better preparation of the examination. They provided the students with factual information about the test and reminded them that their final examination was not very far away. This finding indicated that the HSC examination did have much influence on the teachers. The observation discovered the evidence of negative washback in all around the classroom environment. The class time, lesson contents, activities, use of materials, teacher's behaviours, and teacher's mode of instruction were all influenced by the HSC public examination in English.

5.2.7 Evidence of Washback from the Classroom Observations

The classroom observation were conducted sequentially at selected times, but they were not done continuously. Thus, it is hard to guarantee that they could capture a comprehensive picture of the teaching behaviours in the classroom. However, the data gathered were still representative in the sense that they recorded and reflected typical events and behaviours of the classroom. Overall, the data set presented in this section is qualitative. The next section presents the quantitative data collected through a questionnaire survey. As presented in Chapter Four, this study adopted a mixed-methods approach to data collection and data analysis. Three complementary methods (i.e., interviews, observation and questionnaires) were utilized, with the aim of getting a deeper and more comprehensive understanding of how the role of the examination operates in the washback phenomenon.

As was previously presented in detail, the qualitative data were supplemented with the survey data. The survey was used, for it was assumed to be best suited for quantifying the qualitative data and providing descriptions and comparisons of patterns of teacher beliefs and behaviors. The instrument would permit the generalizability of insights derived from the qualitative data and help the present researcher determines whether the patterns and themes that had emerged

from previous stages could be confirmed and applied to a larger group of participants (questionnaire participants).The Observation-study results reflected that the participants, guided by their personal beliefs, were split in their perceptions of the HSc exam, its impact, and the syllabus and curriculum. Worthy of note is that only two of the six participants (T5, and T7) saw the EFL exam in a positive light. One teacher (T4) suggested that the examination and marking systems should be changed. She also added, listening and speaking should be practised to some extent in the form of IELTS or TOFFL format. While two teachers (T2, T3) expressed negative feelings toward the HSC examination and its impact, their feelings seemed to be mixed.

Some teachers claimed that the test affected their teaching negatively, and asserted that it had a beneficial impact on learning in that it motivated their students to learn. Interestingly, T8 showed negative attitudes towards the HSC examination, and assumed that the examination constrained learning more than it did teaching. Three teachers (T6.T9, T10) commented that the pass rate and number of Grade Point Average–5 (GPA-5) marked the position of their college. In addition, T10 disclosed some crucial points that there was no difference between EFL classes taken in colleges and examination preparation classes arranged in private coaching centres in term of contents of teaching. He complained that many of the students did not attend the classes rather than attended the coaching centres because the examination preparation took place more extensively in the coaching classes. A teacher (T1) viewed, "Some of my students are very irregular in the college, but hardly miss any private coaching class with me at my house".

Through the classroom observation, the present researcher tried to draw a true picture of what happened in the language class for the preparation for the HSC examination. Specifically, the classroom observations convincingly revealed the negative washback both overtly and covertly as Prodomou (1995) delineated. The teachers were found using examples from textbooks that primarily emphasized the skills used in taking the HSC examination. As a result, writing was given much more emphasis in the classroom than listening, speaking, and reading.

5.3 Findings of the Examination Related Documents Analyses

Analyses of examination related documents are crucial to this study because they highlight the problems and characteristics within EFL education and are related with HSC examination. In this study, the present researcher performed analyses of the examination related documents pertaining to the HSC syllabus and curriculum, textbooks used at this level, HSC examination papers, and answer scripts of HSC examination in English. The key purpose of the analyses was to find out what the HSC examination in English set out to measure (e.g., linguistic knowledge or language use) and whether or not the HSC examination represented the curriculum. The analyses also aimed at identifying the characteristics of the HSC examination, for they would serve as the basis for a comparison with what was happening in the classroom, and would help determine whether the observed classroom phenomenon was closely test-related (e.g., whether they were similar or there were gaps between the two).

In this section, the researcher presents and discusses the findings step by step. First, the findings of the syllabus and curriculum analysis are presented. Then, the findings derived from textbooks analysis are reported and discussed. Next, the findings resulted from HSC test (question papers) analysis are documented. Finally, the findings of the HSC answer scripts analysis are presented with discussion. Through the discussion, the present researcher highlights the evidence of washback of the HSC examination on teaching and learning English at the HSC level.

5.3.1 Analysis of the Syllabus and Curriculum

A curriculum should focus on "learners, the subject matter, and society" (Gunter, Estes & Schwab, 2003, p. 14). The authority should: (a) set goals and rational for instruction, (b) define the objectives, (c) decide on means of assessment, (d) construct a breakdown of units of study for the course, and (e) create lesson plans using various instructional models and activities (Gunter et al., 2003). Curriculum developers require information on (a) the needs of the students, (b) the societal purpose [of the learning institution], and (c) the subject matter" (Gunter, Estes & Schwab, 2003, p. 3). Similarly, student needs assessments could provide

background knowledge for teachers prior to planning new learning activities. In addition, teachers may need assistance on how to implement the curriculum so that the content and goals of the lessons align with the standards set by the curriculum. Finally, evaluating the effectiveness of a curriculum program requires authentic assessment of student performance-based tasks (Wiggins, 1997) as demonstrated in the new English curriculum developed in 2000.

Willis (1996) offers five principles of syllabus goals. These provide input, use, reflection on the input and use, and some attention to affect:

1. There should be exposure to worthwhile and authentic language.
2. There should be use of language.
3. Tasks should motivate learners to engage in language use.
4. There should be a focus on language at some points in a task cycle.
5. The focus on language should be more and less prominent at different times.

When considering the syllabus ("... a framework within which activities can be carried out: a teaching device to facilitate learning" Nunan, 1988), this focus leads to specific interpretations of syllabus-design issues as described by Breen and Candlin (1980).

1. What communicative knowledge - and its affective aspects - does the learner already possess and exploit?
2. What communicative abilities - and the skills which manifest them - does the learner already activate and depend upon in using and selecting from his/her established repertoire?
3. Can the performance repertoire of the learner's first language be employed?
4. Can existing knowledge of and about the target repertoire be used?
5. What is the learner's own view of the nature of language?
6. What is the learner's view of learning a language?
7. How does the learner define his/her own learning needs?
8. What is likely to interest the learner both within the target repertoire and the learning process?
9. What are the learner's motivations for learning the target repertoire?

Bangladesh education system is characterised as being examination-driven. One typical example is that students have to sit for numerous examinations as soon as they start schooling. Under this system, examinations are of exaggerated importance. The present curriculum for the HSC EFL education was introduced in 2000, following by the issuance of the new textbooks to be used by the students from 2001.

The National Curriculum and Textbook Board (NCTB) claims that the new syllabus and curriculum at the HSC level follows the communicative approach to teaching and learning English in Bangladesh situation. The NCTB assures that the textbook materials have been designed and developed in such a manner that it can ensure practice in four basic skills of English language: listening, speaking, reading, and writing. As a result, classes are expected to be interactive with students' active participation in the classroom activities through pair work, group work, and individual work. The present HSC English curriculum is considered to be a frontloaded one. The whole syllabus of the English curriculum is accommodated in the textbooks. Two textbooks are prescribed by NCTB for HSC EFL education. *English for Today* for classes 11-12 is considered the mother textbook which was first publish in 2001 while *English Grammar and Composition* is introduced in 2007 as a complementary book to teach grammar as the title implies.

The new frontloaded curriculum, formulated nearly a decade ago, is based on the communicative approach to teaching English, which emphasises students' communicative competence. The English curriculum desires to prepare students for real-life situations in which they may be required to use English. The selection of the course contents has been determined in the light of students' present and future academic, social, and professional needs. The overall aims of the HSC English curriculum (2000) are: to enable the learner to communicate effectively and appropriately in real-life situations; to use English effectively, to develop and integrate the use of the four skills of language, i.e. listening, speaking, reading, and writing; to develop an interest in and appreciation of literature, and to recycle and reinforce structures already learned. A high-stakes test such as the HSC examination influences teaching and learning. Teachers teach those items and skills that are most likely to be tested in the examination. In this situation, it is strongly

believed that communicative teaching or communicative competence is hardly attainable until communicative competence is tested in the examination.

In the syllabus and curriculum of 2000, there was no provision of practising the isolated grammar items. The grammar was supposed to be taught integratedly in discourse and in communication. But the teachers of English were facing challenges teaching English communicatively to attain the desired goals of the EFL curriculum. Though a more communicative competence-oriented curriculum was introduced at the HSC level, the teachers could not shift enough focus from teaching grammar knowledge towards the communicative competences.

The NCTB promised to formulate a guideline for the English teachers, but the guideline did not come into being till today. Since most of the teachers do not have any training to teach at the higher secondary level in communicative approach; since the teachers are to handle big size of classes; since there are almost no facilities of using modern technologies and equipments in the language class; and since there are very limited opportunities for the students to practise speaking and listening inside (due to teacher-cantered classroom) and out side the classroom, the students are found very weak in language form and structure. It is observed that the students remain very weak especially in the formation of new words and sentences both in written and spoken English. In 2007, the government revised the EFL curriculum and introduced grammar and composition items in English second paper. Under the curriculum, a textbook, *Grammar and Composition* was written.

Testing is an integral part of any curriculum. All formal syllabuses make provision of assessing how much of the syllabus is taught, how much the learners have learned, and how much the curriculum objectives are achieved.

5.3.1.1 Findings of the Syllabus and Curriculum Analysis

The study found that the syllabus and curriculum provide ample opportunities for students to use English for a variety of purposes in interesting situations. The emphasis on the communicative approach, however, does not disregard the role of grammar. Instead of treating grammar as a set of rules to be memorised in isolation, the syllabus has integrated grammar items into the lesson activities allowing grammar to assume a more meaningful role in the learning of

English. Thus, students can develop their language skills by practicing language activities, and not merely by knowing the rules of the language. The present English curriculum cannot be separated from the textbooks (prescribed by NCTB) because textbooks represent the curriculum. *English for Today* (for classes 11-12) accommodated all the contents of the syllabus and curriculum. An expert team trained in the UK wrote the book. It was considered a well-suited textbook for practicing EFL at the HSC level. It is also considered the mother textbook for the HSC students.

In keeping with the communicative language teaching (CLT) principles, the English syllabus includes topics of both national and global context, appropriate and interesting to the learners thematically, culturally and linguistically. Adequate grammar contents have also been integrated with language skills so that the elements taught and learned in situations can easily be related to real life situation not just to be memorised as discrete items. It is expected that if used properly, the present syllabus may facilitate learning English through various enjoyable skill practice activities. It provides learners with a variety of materials, such as reading texts, dialogues, pictures, diagrams, tasks and activities; learners can practise language skills using those materials. They can actively participate in pair work, group work or individual work. The syllabus also includes a wide range of topics from both national and global contexts. A curriculum is a vital part of TEFL classes. It provides a focus for the class and sets goals for the students throughout their study. A curriculum also gives the student a guide and idea to what he/she will learn, and how he/she has progressed when the course is over. The test leads to the narrowing of contents in the curriculum.

The analysis of the syllabus and curriculum finds that the HSC syllabus and curriculum is communicative thematically, but there is a very question whether the set objectives of the curriculum are attainable. Because the teachers do not like to take any risk of teaching the items which are not tested, they consider it simply waste of time, they skip items and narrow down the syllabus and curriculum contents towards the preparation of the examination. The present study found that both the teachers and the students were very selective in choosing study contents for the preparation of examination. That is, teachers design the classroom activities as per the test contents. This practice is an evidence of washback effect on the syllabus

and curriculum. The present HSC English syllabus and curriculum do not affect the test or teaching, but HSC examination affects the syllabus and curriculum.

The present English curriculum was influenced by research in the fields of foreign language learning, education, assessment, cognitive psychology and curriculum development. The principles underlying: (a) language learning and teaching, (b) choice of materials, content, and tasks, (c) classroom assessment; formative and summative, (d) alternatives in assessment, and assessment requirements and criteria, and (e) the role of the pupils align with a constructivist approach to curriculum development and learning (Posner, 2004). In addition, the principles underlying language teaching also follow brain-based learning theories that cater to learners' needs; preference for learning styles and multiple intelligences. The English curriculum artifact provides teachers and learners with a constructivist approach to assessment "as an integral part of the teaching-learning process with guidelines and on expectations for formative and summative assessments, and criteria for alternative assessments that would reflect performance in the target language competencies described in the curriculum.

The new English curriculum is a well planned EFL artifact that enhances student performance and embraces different learning styles (Rabbe & Shuster-Bouskila, 2001) by supporting brained-based learning.

5.3.1.2. Evidence of Washback on the Syllabus and Curriculum

Testing is a vibrant art of the curriculum; the test contents and items should be determined in line with the objectives of syllabus and curriculum. Since the teachers are the main stakeholders to implement the agendas of the syllabus and curriculum, they should have been given a set of guidelines to follow for achieving the targets. If the examination system does not test communicative competence of the students or the four skills of language, the teachers will not teach other skills which are unlikely to be tested in examination. The findings of the questionnaire surveys revealed that 64% students and 59% teachers confirmed that they were not aware of the objectives of syllabus and curriculum. It was also found that 74% students and 64% teachers believed that the present syllabus could enhance EFL teaching and learning. The present study also revealed that 86% teachers and 72%

students did not care about the syllabus and curriculum because they practised what were important for the examination. During the survey, 60% teachers and 71% students pointed out that they did not practise all the sections and contents of syllabus and curriculum. The findings of classroom observation and interview with the teachers also revealed that they did not teach the syllabus, rather they taught to the test.

It was also found that both teachers and students preferred to use test related commercially produced materials such as guidebook, suggestion book, model test papers, etc. The classroom observation found that over 80% teachers taught to the test directly and heavily dependent on the commercially produced materials. These test preparation materials are termed as 'hidden syllabuses by many researchers (e.g. Caine, 2005 and Wang, 2010). This is the powerful evidence of existing negative washback of the HSC examination on the EFL teaching and learning in general and on the syllabus and curriculum in particular.

In itself, however, any syllabus and curriculum cannot ensure that communicative language teaching and learning take place in the classroom. It can only provide a set of criteria which, if properly implemented, would give the best possible change for that to happen. The present HSC examination influences the teachers to teach to the test opposed to teach the syllabus. Test contents also can have a very direct washback effect upon teaching curricula. Therefore, curriculum is a vital part of the EFL classes. Very often the test leads to the narrowing of contents in the curriculum. Alderson & Wall (1993) point out that test can affect curriculum and learning.

It is believed that washback has deep relation with the syllabus and curriculum. Frontloading alignment of curriculum is commonly practiced in EFL education. A frontloaded curriculum can prevent teaching to the test, which may lead to an extremely narrow and rigid view of the actual goals and objectives of any curriculum. The findings from the study about washback onto the curriculum indicate that it operates in different ways in different situations.

5.3.2 Textbook Material Analysis

Textbook materials play a very important role in language classrooms. .A textbook is a tool, and the teacher must know not only how to use it, but how useful it can be. The purpose of the text book analysis was to determine the overall pedagogical value and suitability of the book towards this specific language programme. In Bangladesh, the EFL teachers and the learners use two types of materials such as textbooks prescribed by the authority, that is, National Curriculum and Textbook Board (NCTB), and commercially produced examination related materials (e.g. guide books, suggestion boos, model test papers, etc). In many contexts, language teachers are heavily reliant on available materials and this is perhaps even more evident in the testing context where teachers may feel that following a test preparation book is the safest way to ensure all the crucial points are covered. As with other high-stakes tests, the HSC examination in English aims to assess students' general level of language ability and is therefore linked to particular materials or programme of instruction. Nevertheless, the majority of teachers in Bangladesh are dependent to a large extent on materials focusing specifically on examination preparation other than textbooks prescribed by NCTB. Bailey (1999) suggests that textbook washback is a possible result of test use. She points out that test preparation materials are the indirect evidence of washback. The appropriateness of a textbook and therefore any consideration of the possible existence of washback must be considered within the specific context in which it is being used.

Shohamy (1992, p.514) states, "… negative washback to programs can result in the narrowing of the curriculum in ways inconsistent with real learning and the real needs of … students". The opinion that there is the potential for texts to narrow the curriculum and encourage negative washback is also reported by Cheng (1997), Shohamy et al. (1996), and Alderson & Hamp-Lyons (1996). The literature provides many references of materials being linked to negative washback both in terms of their content and their classroom use. The use of these kinds of materials in classrooms has an effect on how the students view test preparation, and how they make ready themselves for the test. Fullilove observes that texts which are "little more than clones of past exam papers" resulted in some students spending time memorising model answers at the expense of learning how to create answers to similar questions (1992, p. 139).

With so much written about the potential of textbooks to have a negative effect on teaching and learning, the question is what features would be desirable in a test preparation text for it to have a positive effect. Referring specifically to High-stakes preparation texts, Hamp-Lyons (1998p. 330) makes the statement that such books should "…support teachers in their principal task of helping learners increase their knowledge of and ability to use English." She identifies some characteristics a textbook having positive washback might require:

> …the inclusion of appropriate content carefully designed to match learning needs and sequence and planned to support good classroom pedagogic practices; it also requires keeping close sight of what is appropriate in test preparation practices and what the demands of the test itself are (ibid: 330).

The effectiveness of commercial test-preparation materials used by way of preparation for standardised tests such as the HSC examination is still under-researched. Such materials may be appropriate depending on how closely they match the test and the inference one wishes to make from the test scores. Investigating washback in the context of the Hong Kong Certificate of Education Examination, Cheng comments:

> We believe teaching and learning should include more varied activities than the examination formats alone. However, it would be natural for teachers to employ activities similar to those activities required in the examination (1999, p. 49).

Lam, (1994) finds that about 50% of the teachers appear to be "textbook slaves" in teaching the sections of the test related items. Cheng (1997, p.50) also notes the existence of workbooks specifically designed to prepare students for examination papers in the Hong Kong Certificate of Education Examination and the heavy reliance of teachers on these workbooks.

On the topic of textbook evaluation, Williams (1983, p.254) highlights the importance of considering the context within which a textbook is used. The fact that test preparation books for the HSC examination can be considered a part of the impact of the test. The development of textbooks which claim to prepare students for an examination can be seen as a kind of evidence of washback. The type of materials they contain and the approach they take can be used as an indication of whether the washback of the examination is positive or not. One feature that one would expect in

a language classroom is the inclusion of input and exercises that explore the components of the language.

This reliance on commercially produced materials in this context is evidence of negative washback because instead of introducing more authentic materials and prescribed textbooks by the authority the teachers prefer to use commercial textbooks, most of which are basically modified copies of the HSC examination paper. The present study evaluated *English for Today* (for classes 11-12) to check whether the book represented the English syllabus and curriculum, and to look into whether the HSC examination in English had any washback (positive or negative) on *English for Today* for classes 11-12. The study did not take any attempt to evaluate *English Grammar and Composition* and the commercially produced materials.

5.3.2.1 Justification for Textbook Evaluation

It is important to remember, however, that since the 1970's there has been a movement to make learners the center of language instruction and it is probably the best to view textbooks as resources in achieving aims and objectives that have already been set in terms of learner needs. Moreover, they should not necessarily determine the aims themselves (components of teaching and learning) or become the aims but they should always be at the service of the teachers and learners (Brown, 1995). Consequently, efforts must be made to establish and apply a wide variety of relevant and contextually appropriate criteria for the evaluation of the textbooks that can be used in language classrooms. It should also be ensured "that careful selection is made, and that the materials selected closely reflect [the needs of the learners and] the aims, methods, and values of the teaching program" (Cunningsworth, 1995, p.7). Sheldon (1988) has offered several other reasons for textbook evaluation. He (ibid.) suggests that the selection of an ELT textbook often signals an important administrative and educational decision in which there is considerable professional, financial, or even political investment.

Moreover, it would provide for a sense of familiarity with a book's content thus assisting educators in identifying the particular strengths and weaknesses in textbooks already in use. This would go a long way in ultimately assisting teachers

with making optimum use of a book's strong points and recognizing the shortcomings of certain exercises, tasks, and entire texts. One additional reason for textbook evaluation is the fact that it can be very useful in teacher development and professional growth. Cunningsworth (1995) and Ellis (1997) suggest that textbook evaluation helps teachers move beyond impressionistic assessments and it helps them to acquire useful, accurate, systematic, and contextual insights into the overall nature of textbook material. Textbook evaluation, therefore, can potentially be a particularly worthwhile means of conducting action research as well as a form of professional empowerment and improvement. Similarly, textbook evaluation can also be a valuable component of teacher training programs for it serves the dual purpose of making student teachers aware of important features to look for in textbooks while familiarizing them with a wide range of published language instruction materials.

5.3.2.2 Textbook Analysis Checklist

ELT materials play a very important role in many language classrooms, but in recent years there has been a lot of debate among the ELT professionals on the actual role of materials in Teaching English as a Second/Foreign Language (TESL/TEFL). Arguments have encompassed both the potentials and the limitations of materials for 'guiding' students through the learning process and curriculum as well as the needs and preferences of teachers who are using textbooks. Other issues that have arisen in recent years include textbook design and practicality, methodological validity, the role of textbooks in innovation, the authenticity of materials in terms of their representation of language, and the appropriateness of gender representation, subject matter, and cultural components.

Although Sheldon (1988) suggests that no general list of criteria can ever really be applied to all teaching and learning contexts without considerable modification, most of these standardised textbook evaluation checklists contain similar components that can be used as helpful starting points for ELT practitioners in a wide variety of situations. Preeminent theorists in the field of ELT textbook design and analysis such as Williams (1983), Sheldon (1988), Brown (1995), Cunningsworth (1995) and Harmer (1996) all agree, for instance, that evaluation

checklists should have some criteria pertaining to the physical characteristics of textbooks such as layout, organisational, and logistical characteristics.

Other important criteria that should be incorporated are those that assess a textbook's methodology, aims, and approaches and the degree to which a set of materials is not only teachable but also fits the needs of the individual teacher's approach as well as the overall curriculum. Moreover, criteria should analyse the specific language, functions, grammar, and skills content that are covered by a particular textbook as well as the relevance of linguistic items to the prevailing socio-cultural environment. Finally, textbook evaluations should include criteria that pertain to representation of cultural and gender components in addition to the extent to which the linguistic items, subjects, content, and topics match up to students' personalities, backgrounds, needs, and interests as well as those of the teacher and/or institution.

The present study evaluated the textbook, *English for Today* for required information for the study. A checklist (Appendix-3A) was applied for the analysis which was adapted from the American Council on the Teaching of Foreign Languages (ACTFL). A number of textbook evaluation checklists and guidelines had also been studied for the present study to evaluate *English for Today* for classes 11-12.

5.3.2.3 Analysis of *English for Today* for Classes 11-12

National Curriculum and Textbook Board (NCTB) prescribed two books: *English for Today* for classes 11-12, and *English Grammar and Composition* to cover the entire HSC English syllabus. One of the curriculum experts at NCTB points out that *English for Today* (for classes 11-12) is considered the "mother textbook" while *English Grammar and Composition* is a complementary book designed for test purpose. *English for Today* (EFT) was written by NCTB in 2000 when communicative approach was introduced at the HSC level. *English Grammar and Composition* was written when English 2nd paper was revised in 2007. *English for Today* (for classes 11-12) was considered a well-suited textbook for practicing EFL at the HSC level.

The source materials in the textbook were taken from authentic materials of everyday life. It was claimed to have been designed to reflect and reinforce the communicative competence in terms of teaching and learning objectives, focuses, and approaches. This textbook differs from that of the past in that more cultural-related themes were incorporated in the content of the materials. It has also been noticed that English teachers in Bangladesh find themselves in an unenviable position in which the constraints imposed by the examination-driven "hidden syllabus" prevent them from implementing, in practice, communicative methodology. Though the textbook was written nearly a decade ago, no revision, inclusion, innovation has been made to it to date. The textbook *English for Today* for classes 11-12 is made of high-grade, durable paper and the presentation of information appears to be clear, concise, and user-friendly. The book also contains many charts, models, and photographs that help clarify and contextualize information while the presence of hand-drawn pictures portrays a friendly and humorous atmosphere.

There is no separate textbook edition for the teacher that could be used as a methodological guide or so called "teacher's edition." The textbook should provide appropriate guidance for the teacher of English. Though the authority intended to formulate a 'guideline' for the teachers on how to use the book, no guideline has been written yet. The untrained, or partially trained teachers who do not possess enough control overall aspects of English should not be left in any doubt concerning the procedures proposed by the textbook. Otherwise, he or she may, for example, teach only the meanings of the minimal pair 'live/leave', completely ignoring the writer's intention that these items should be used for pronunciation practice.

The EFL textbook, should give introductory guidance on the presentation of language items and skills. The textbook serves as a syllabus. The carefully planned and balanced selections of language contents enable teachers and students to follow subject systematically. The course book can provide useful guidance and support particularly for teacher who are inexperienced. It suggests aids for the teaching of pronunciation (e.g. phonetic system), offer meaningful situations and a variety of techniques for teaching structural units, distinguish the different purposes and skills involved in the teaching of vocabulary, provide guidance on the initial presentation of passages for reading comprehension, demonstrate the various devices for

controlling and guiding content and expression in composition exercises, contain appropriate pictures, diagrams, tables, etc.

The textbook, *English for Today* (for classes 11-12) represents the HSC English syllabus and curriculum. The learners and the authors of the textbook belong to same linguistic background. The writers of *English for Today* (for classes 11-12) are   the people of Bangladeshi, but they are highly experienced in English language teaching and trained in the UK. The book contains 24 units comprising 156 lessons; every lesson has a set of objectives. Most of the lessons outline new theme and task. Almost every lesson contains exercises that may promote language skills. Yet, listening exercises are hardly incorporated. *English for Today* is mostly student-centred. There are lots of activities, tasks, exercises where students' participations are must such as pair work, group work, individual work, making dialogues, amplifying ideas, making answer to questions, etc. In all the activities, the learners have to comprehend and/or produce language, i.e., they have to use language, "do the exercises either individually" or "in pairs or in groups". For example, Unit One: lesson 1, E (page 3); Unit Five: Lesson 2, E (p. 61). Most of the tasks of the lessons are enjoyable.

The textbook *English for Today* includes a good number of stories and articles on social affairs, historical events, educational subjects, wonders, heritage, space, communication, challenges, profession, sports issues, etc. (such as caring and sharing, email, looking for a job, etc); therefore, it may be termed as well-suited and interesting one. Many of the lessons and topics are interesting thematically and conceptually. But the presentation of the tasks and activities are stereotypical and traditional because the lessons start with a typical activity (e.g. looking at the picture(s)). When most of the lessons start with such types of stereotypical activities, learners as well as the teachers get in difficulty to carry out them. They may feel bore. For example, Unit One: Lesson 1 'Our Family' (A) -Look at the picture of Nazneen's family; Unit One, Lesson 2 'A Myanmar family' (A) -Look at the picture below and exchange your views with your partners); Unit One: Lesson 4 'Mr. Fraser's family' (Look at the picture what kind of person do you think he is? Discuss in pairs).

In the present textbook, "Look at the picture(s)" is presented in most of the lessons; it is in the beginning or somewhere else. Although some pictures are

considerably different from others in terms of physical contexts, students are not provided with any linguistic context at the beginning. As a result, these may often produce boredom among the pupils, and teachers may face difficulty to arousing interest among the learners.

The instructions given in *English for Today* (EFT) are clear and easy to understand for the learners. Even if, the learners might not be familiar with the structures and the lexis used in the instructions, the models given for each group of exercises provide contextual clues for the learners as to what they are expected to do. However, some of the instructions lack the required contextual information in terms of linguistic contextual complexity. For example, Unit One: Lesson-A (p. 9)

The textbook may be considered appropriate for the HSC level students in Bangladesh context. The book maintains difficulty level at the 12th grade standard in respect of text and exercises. *English for Today* (for classes 11-12) does not include any topic on explicit grammar practice. Implicit grammar is presented thoroughly in different items. There is no scope of traditional grammar practices in the lesson, rather, grammar items and their functions are included within the text and discourse of varied types in each lesson in the implicit manner. This point has been made clear in the book map of the book.

Each and every lesson presents implicit grammatical exercises, such as tenses, clauses, verbs, comparison, modals, direct and indirect speech, change of voices are presented in the lessons through various exercises i.e. identification, right form of verb, fill in the gaps with clues, fill in the gaps without clues, matching column, etc. The grammar comes into different tasks and activities, but not in an isolated manner in any way. For examples, (1) "Match the verb in column A with the definition in column B" (Unit Six: Lesson 4, Page 77), and (2) "Use the appropriate forms of the given words to complete the following sentences" (Lesson 3: Unit Six Question-C, p.75). Lessons indicate what students should know, and be able to do. Increase order of difficulty is maintained. Almost every lesson contains comprehension exercises, grammar, etc. But the ideas are not sequential. The students are given some guidelines to perform tasks such as page 3, task E.

Maximum learners, taking little help, can use the textbook on their own. New vocabulary items are presented in a table at the end of every lesson. Each and every lesson provides scopes for practicing vocabulary through different techniques.

Vocabulary items are explained through defining the word or and providing synonyms. Repetitions of vocabulary items are hardly found; new vocabulary items come up to be practiced in different lessons and exercises. There are ample opportunities for practicing dialogues, but *English for Today* (for classes 11-12) does encourage neither the teachers nor the students to use audio/tape recorder or any audio – visual aids. Not enough illustrations, charts, etc. are used. Sometimes, the pictures do not relate to the idea that a sentence is showing. For example, in page 4 (Unit One: lesson 2) the picture does not represent the idea of the lesson. No separate printed material is provided in this textbook, but the textbook has used many authentic contents.

Nearly 80% textbook contents are realistic, taken from everyday social life. Social environments are represented in the textbook; no religious belief and environment are represented. In some cases, there is a connection between the previous lessons, for example, page 294, "The challenge ahead- I" and (p. 295), "The challenge ahead- II". These lessons are related – introducing the challenges ahead, but not all lessons are related with the previous or the next one. In many cases, the title of the lesson does not indicate the aim of the lesson, for example, Unit Twenty-two: Lesson 4, and Unit Fourteen: Lesson 4.

There are ample opportunities in the book to use the target language such as dialogues, pair work, group discussion, etc. The lesson describes sequence of instructional activities, and assessment. This textbook does not provide methodology guide for the teacher. Learners' native language is discouraged in the English class though limited use of the first language is allowed. Culturally known lessons create interest among the students; therefore, lessons should be relevant to the day to day activities of the learner. Many topics of the textbook are taken from the natives' cultural, social, educational and historical background, though some lessons are extracted from students' unknown arena of subject.

A few contents in *English for Today* are not fit for the students because they do not connect to a certain degree of reality, for example, the lesson "The London Underground" (Unit 16, Lesson 7, p. 209) which is devoid of reality in Bangladesh context. There are hardly any instructions in this entire textbook, but examples are adequately explained and illustrated for the students. It is possible to set up groupings varied in response to the nature of learning. Most of the lessons are

relevant to learners' life and culture. The learning opportunities in the textbook are real and rich in ways that promote students' engagement and interest. There are a good number of activities for the students to apply their knowledge to practical and real-world situations. There are adequate opportunities for the students (such as dialogues) to be creative in their day to day correspondence.

The book contains lessons on modern mode of communication such as faxes, emails. For example, Unit Seventeen: Lesson 5, 'Fax' (p.219), and lesson 6, 'E-mail'. *English for Today* also contains a unit on 'Conquering Space' (Unit Twenty-two, p.277). The textbook itself does not emphasize any lessons or tasks for test purpose. The opportunity of self assessment is limited. The textbook is good enough for learning English as a foreign language. No guidelines have been provided for the examination preparation, except the syllabus and marks distribution in the preliminary page section. The textbook *English for Today* does not discuss the examination, and provide any examination tips for the students.

5.3.2.4 Findings of *English for Today* Analysis

The Textbook "*English for Today*" for classes 11-12 takes into account currently accepted methods of EFL teaching. It gives guidance in the presentation of language items. The book relates content to the learners' culture and environment. It includes interesting contents to good extent. It suggests ways of demonstrating and practising speech items. The book includes speech situations relevant to the pupils' background. It allows for variation in the accents of non-native speakers of English.

English for Today stresses communicative competence in teaching structural items. It provides adequate models featuring the structures to be taught. The book clearly shows the kinds of responses required in drills (e.g. substitution). It selects vocabulary items on the basis of frequency, functional load, etc. It distinguishes between receptive and productive skills in vocabulary teaching. The book presents vocabulary items in appropriate contexts and situations. The book focuses on problems of usage related to social background.

It does not offer exclusive listening exercises. It includes teachers' speech, explanations, dialogues, pair work, etc. for practising integrated skills. The book offers no instruments and equipment for practising listening. *English for Today*

includes dialogues and discussions. It offers exercises on asking questions. It presents integrated skills practice exercises; offers story telling opportunities. *English for Today* offers exercises for understanding of plain sense and implied meaning, relates reading passages to the learners' background, selects passages within the vocabulary range of the pupils and selects passages reflecting a variety of styles of contemporary English. *English for Today* relates written work to structures and vocabulary items practised orally. It gives practice in controlled and guided composition in the early stages. The book relates written work to the pupils' age, interests, and environment. It demonstrates techniques for handling aspects of composition teaching.

Potential of textbooks to create washback is well documented in the literature. Key issues in 'textbook washback' include the role that publishers and authors play in influencing the types of preparation materials that come onto the market, and the role of teachers as the interpreters and presenters of the contents of the books. The features seen as promoting positive washback in textbooks follow on from the literature in general, with the importance of including not only information about the requirements of the test, but also tasks that support good classroom.

As textbooks are the primary source of classroom materials, their content and approach have a direct impact on what happened in the classrooms. It is important to realize that the teaching materials selected by teachers vary from class to class. In general, there are four major types of materials used in the observed EFL classes: *English for Today* (EFT), the HSC test papers, guidebook, and HSC model questions. HSC examination-related materials concern those materials used for fostering students' test-taking strategies. The HSC test papers here refer to the printed books of question papers previously used in the HSC examination and in the model examinations in different renowned colleges.

It is worth keeping in mind that it is a common practice among the Bangladeshi EFL teachers and students to use more than HSC test papers prior to the final examination. The key reason is that they want to use the papers to familiarize their students with the test format. These materials are not authorized by the government; they are commercially produced for the purpose of test preparation. During classroom observation the researcher found 7 teachers (out of 10) practising test paper, guidebook and past questions in the class. They did not bring

the original textbook (EFT) with them for classroom use. They claimed that supplementary materials reflect the test contents in their objectives, emphasis and approach, and to reinforce general goals of test preparation. These supplementary materials are termed as hidden syllabus by many researchers (e.g. Caine, 2005; Wang, 2010).

The book is also very attractive and organized in a clear, logical, and coherent manner. This organization reflects a topic-based structural-functional syllabus that is designed with the goal of facilitating communicative competence. In addition, the activities and tasks in *English for Today* were found to be basically communicative and they seemed to consistently promote a balance of activities approach. This in turn encouraged both controlled practice with language skills as well as creative, personal, and free responses on the part of the students. Despite its strengths, *EFT* still had some shortcomings. *English for Today* for classes 11-12 can be an effective textbook in the hands of a good teacher and instructors should not be discouraged from using it with the appropriate learner audience.

5.3.3 Analysis of the HSC English Test

Testing is universally accepted as an integral part of teaching and learning. It is one of the basic components of any curriculum, and plays a pivotal role in determining what learners learn. Candlin and Edelhoff (1982, p.vi) assert that 'learners learn most when they are quite precisely aware of . . . how their efforts are to be judged and evaluated'. Testing also plays a central role in deciding what teachers teach and how they teach; Reardon et al. (1994), for example, contend that 'changes in assessment policies can be used as a powerful lever for reforming schools'. Bailey (1996, p.276) suggests, "… a test will yield positive washback to the learner and to the programme to the extent that it utilises authentic tasks and authentic texts".

Both teachers and students have attitudes to tailoring their classroom activities to the demands of the test, especially when the test is very important to the future of the students, and pass rates are used as a measure of teacher success. This influence of the test on the classroom is, of course, very important; this washback effect can be either beneficial or harmful. The framework proposed by Bachman and

Palmer (1996) is often taken as a theoretically grounded guideline (Appendix-3E) for analysing the characteristics of a test. This conceptual framework consists of a set of principles involving five facets of tasks: setting, test rubric, input, expected response, and relationship between input and response. But here, the present researcher presents and discusses four features in particular which he thinks are crucial for this study.

The present HSC English syllabus and curriculum introduced communicative approach of teaching and learning, and set a number of goals and objectives to be achieved. A test is a part of curriculum, so, the test should reflect and correspond to the syllabus and curriculum. The present study performed the HSC test (English *First Paper* and *Second Paper*) analysis to examine the nature, contents, characteristics, and their influence (washback) on classroom teaching and learning. The present researcher visited the concepts of each facet and analysed the HSC examination in light of the features of this framework.

5.3.3.1 Task Characteristics and Contents

Test contents can have a very direct positive or negative washback upon teaching curricula and classroom behaviour. As Pearson remarked, 'There is an explicit intention to use tests, including public examinations, as levers which will persuade teachers and learners to pay serious attention to communicative skills and to teaching learning activities that are more likely to be helpful in the development of such skills (1988). First, the test rubric facets are portrayed by Bachman and Palmer (1996) as those characteristics of the test that provide the structure for a particular test. These characteristics include instructions and the structure of the test (e.g., how the test is organised), time allotment (e.g., the duration of the test as a whole and of the individual tasks), and scoring procedures (e.g., how the language that is used will be evaluated, or scored).

A close look at the HSC examination in English indicates that its instructions are not clearly specified with respect to the procedures for taking the test. No examples are provided as to how to perform a task. For instance, in the *First Paper* (question-5), the section of 'Seen Comprehension' (Skimming and Scanning), the test takers are not provided with explicit directions except being told "to answer the

questions" (Appendix-3C). Further, little information is given on score distribution, criteria for correctness, and procedures for scoring the responses (e.g., how the test will be scored as well as the steps involved in scoring the test).

It should be noted that writing skills and grammar items still take up a larger percentage of the test. The each part of the question paper (*First Paper*) includes different types of tasks and activities. Though the syllabus is communicative in nature; the textbook has been written with communicative view of teaching and learning, the students are not tested communicatively; that is, communicative competence of the examinees is not tested in the examination. No tasks and items are designed in the examination to test listening and speaking skills.

English *First Paper* comprises 3 parts (Appendix -3C). The questions from 1-4 are related with reading comprehension along with writing skills allocating 20 marks; the questions from 5 -8 aim at testing student's writing ability covering 20 marks; the questions 9 -10 are vocabulary and grammar items carrying 20 marks; and the questions 11-13 test students inductive knowledge and writing skills. As it is shown in the table (Table 5.73), the tasks in Part-A, Part-B, and Part-C mainly measure reading comprehension, writing, vocabulary, and grammar. The Part- C (questions 11, 12 and 13) carrying 40 marks is an observable evidence of invalid language test. Those items do not maintain the difficulty level of HSC standard. Matching phrases, re-arranging sentences, writing a paragraph (of 150 words) by answering questions have been given exaggerated importance by allocating excessive marks.

Bachman (2000) reports contents of a test must include activities that gauge students language ability if it were communicative language testing. He adds that test objectives should reciprocate curricular objectives. Andrews (2004) points out that when the test does not match the curriculum objectives, it generates negative washback on teaching and learning. During the classroom observation, the researcher observed that most of the teachers (nearly 80%) directly teach to the test. Furthermore, the findings can be cross-referenced with the findings derived from questionnaire survey where the researcher found that teachers practised those tasks and activities which are tested in the examination. According to Bachman and Palmer (1996), these task types are limited in that they may not be appropriate for measuring different areas of language ability. The table below (Table 5.73) shows

that the contents and tasks of the test have followed the nature and types of exercises of *English for Today*, but it can only test learners reading and writing ability:

Table 5.73: Test contents and marks distribution –*First Paper*

English First Paper : Total Marks-100			Time -3.00 Hours		Skill and element tested
Part-A	Q. No	**Seen Comprehension**: **Marks- 40** (Skimming and Scanning)	No. of tasks	Marks	
	1	Choosing right word to complete sentences	5	5	-Reading -Comprehension -Writing
	2	True/false, if false, providing right answer.	5	5	-Reading comprehension, writing,
	3	Filling the gaps with correct form of words in brackets, in a given sentence (with clues).	5	5	-Reading -Comprehension, - Language form - Writing
	4	More free/open questions (knowledge based)	5	5	-Reading Comprehension Writing
	5	Short open questions	5	5	Writing
	6	Filling the gaps in a given discourse (without clues).	5		Writing
	7	Summarising a given passage	1	5	Writing
	8	Making short notes (in flow -chart)	5	5	Writing
Part-B		**Vocabulary: Marks- 20**			
	9	Filling the gaps (Cloze test with clues)	5	10	-Vocabulary -Grammar
	10	Filling the gaps (Cloze test without clues)	10	10	-Vocabulary -Grammar
Part-C		**Guided Writing: Marks-40**			
	11	Matching the phrases to make sentences	6	12	-Contextual knowledge -Writing
	12	Re-arranging the jumbled sentences	14	14	-Contextual knowledge -Writing
	13	Writing a paragraph answering a set of questions	1	14	-Writing

English *Second Paper* (Appendix-3D) comprises two parts: grammar, and composition. In this paper of English, there are no scope of testing listening, speaking, and reading. Three questions (1-3) in Part -A carry 15 marks on the use of grammar inductively. It is good to observe that there is no implication of isolated or discreet-point of grammar testing. The question 4 is on sentence making, which

323

absolutely tests writing skills of the students. The details of marks distribution are illustrated in the Table 5.74:

Table 5.74: Test contents and distribution of marks- *Second Paper*

English Second Paper: Total Marks-100					Skill and element tested
Part-A	**No**	**Grammar : Marks-40**	No. of tasks	**Marks**	
	1	Using suitable verb (from a list) in a given passage	5	5	Grammar (Integrative)
	2	Filling in the blanks with suitable preposition in a given discourses:	5	5	Grammar (Integrative)
	3	Use of articles where necessary.	10	5	Grammar (Integrative)
	4	Making meaningful sentences with idioms.	5	5	Writing
	5	Change of speech	5	5	Grammar
	6	Transformation of sentences	5	5	Grammar
	7	Tag questions	5	5	Grammar
	8	Completing sentences using conjunction.	5	5	Grammar
Part-B	**Composition**				60 marks
	9	Report writing	1	10	-Writing (Subjective)
	10	Short composition/paragraph	1	15	-Writing (Subjective)
	11	Letter writing	1	10	Writing (Subjective)
	12	Writing a dialogue or writing a summary of a given passage	1	10	Writing (Subjective)
	13	Completing a story	1	15	

The questions from 5-8, covering 20 marks, are related with the grammar items, but they are less integrated and less inductive. The Part-B consists of 5 questions on comprehension that carries 60 marks. The questions of this part are subjective rather than objective. The questions from 9-13 assess students' grammar, vocabulary, and inductive knowledge. English *Second Paper* is more writing skills oriented than *First Paper*. Due to nature of tasks and contents, S*econd Paper* has no opportunity of testing listening, speaking, and reading. Reliability and validity of the examination in terms of coverage of the curriculum, and objectives of the syllabus are considered dubious. The present HSC English curriculum introduced communicative approach of teaching and learning. It was expected that the new textbook would facilitate communicative competence. The English syllabus and curriculum claimed that lot of communicative exercises had been incorporated, and

324

therefore practice of the four skills of English would be geared up. The syllabus of English *Second Paper* was revised in 2007, and the first examination under the new syllabus was held in 2009. In the new syllabus, more grammar and composition items were introduced. Then, a Textbook, *English Grammar and Composition* was prescribed to meet the demand of the test.

A crucial change of the HSC examination from the previous tests was testing students' communication competence rather than grammar competence. Over 80% of the interviewed teachers indicated that the contents or format of the earlier version (before 2009) of the HSC *Second Paper* in English was more real-life oriented compared to the present version. Thus, their teaching was shifted from promoting their communication ability toward drilling students' grammar knowledge. The test tasks do not often correspond to tasks in the language use (TLU) domain. In other words, the content of the test does not include many items related to daily life situations. Of the small number of tasks, writing is the only one that represents a real-life task. Apart from it, other test tasks, both the tasks in multiple type format and those in constructed response formats can hardly relate to every day situations. In other words, they do not assess students' ability to use language through real-life situations. Bachman and Palmer (1996) describe that interactive and conversational tasks involve a give and take.

The above analysis also reveals that the revised HSC examination in English does not assess test takers' overall language skills integratively. As mentioned above, a test taker's linguistic competence is adequately assessed, for the majority of test items only test his or her areas of language knowledge. Apart from writing, other test tasks of the test cannot be interpreted as assessing test takers' sociolinguistic competence or strategic competence. In the section of Composition, a testee's language competencies are integratively tested, for he or she is expected to produce language in the same way as in everyday contexts. In addition to linguistic knowledge, the task involves other areas of language knowledge as well as metacognitive strategies. For instance, the language the testee produces must not only be accurate (evidence of linguistic competence) but also appropriate (evidence of sociolinguistic competence) and coherent (evidence of discourse competence).

One drawback of the HSC examination in English for both first paper and second paper is that the test does not include an assessment of students' ability to

perform listening and speaking tasks. The testees do not have the opportunity to demonstrate strategic competence in the oral interview. In this regard, although the curriculum is claimed to assess students' ability to use language through real-life tasks, its results or test scores cannot truly reflect test takers' communicative competence. Owing to the fact that little evidence can be retrieved to show that the HSC test score reflects the area(s) of language ability it sets out to measure, its construct validity is called into question.

5.3.3.2 Input

The input facets examine the format in which input is presented and the characteristics of the language that are embodied in the input. The format includes features such as channel, form, language, length, type, degree of 'speededness', and vehicle of input delivery, while the language characteristics include both organizational and pragmatic characteristics of how the language is organized.

With respect to the HSC examination, the form of input is language, and the language by means of which input is delivered is English. On the whole, the input is presented in extended discourse. Since the majority of the task types in the HSC examination are objective or right forms of verbs/words, short questions, vocabulary, matching, true/false, cloze test and composition, etc. The degree of 'speededness' is high, since the rate at which the test taker is expected to process the input information is high. As a test for measuring teaching results, the HSC examination in English is an achievement test.

Although it refers to the syllabus, it seldom takes teaching contents into consideration. This causes the separation of tests from teaching goals, which, in turn, causes students to value tests more than regular class performance. Many students think that so long as they can pass the test it does not matter whether they attend the regular classes or not, which results in students' high rate of absence from classes in some colleges. Therefore, by focusing on testing students' reading comprehension and writing ability while neglecting their productive ability, the present HSC English test cannot reflect students' communicative competence objectively, and thus its validity is doubtful.

5.3.3.3 The Nature of Language Input of the HSC Examination

The organizational and pragmatic characteristics of the HSC examination both for *First Paper* and *Second Paper* can be divided into subcategories. As far as the HSC examination is concerned, it involves a broad range of vocabulary and grammatical structures as well as a wide range of cohesive devices and topics. An examination of the test shows that a testee's linguistic knowledge is adequately tested, because he or she has to demonstrate that he or she has the linguistic knowledge to process the input information.

However, testees' sociolinguistic competence is only tested to some extent. For instance, in Part –A of *First paper*, the section of 'Seen Comprehension', and this competence is tested, for a testee needs to guess word meanings in context. But in general, the language input of the HSC examination seems unnatural, for it seldom relates to everyday situations. In other words, the test tasks do not often correspond to tasks in the language use domain. Based on the view of Bachman and Palmer (1996), if we want to use the scores from a language test to make inferences about individuals' language ability, we must be able to demonstrate how performance on that language test is related to language use in specific situations other than the language test itself.

5.3.3.4 Validly and Reliability of the HSC EFL Test

The validity of the HSC examination in English is doubtful. The study indicates that HSC test scores, with a questionable reliability and doubtful validity, can not objectively reflect the students' normal English level, and its negative washback effect has hindered students' development in their communicative competence. The analyses of the examination papers support the findings of classroom observation and questionnaire survey. The classroom observation of this study found that teachers did not practice listening and speaking in the class at all. The reason for not teaching those skills (e.g. listening, speaking) was that they were not tested. Furthermore, the findings resulted from the questionnaire surveys reported that listening and speaking were not practised in the class. In addition, teachers did not teach items and tasks which were not tested; they considered it simply waste of time. In the questionnaire survey, both teachers and students pointed

out that the present HSC examination in English hindered their teaching and learning English. The results of the interview with the EFL teachers revealed that the present examination system obstructed their teaching. The findings also showed that maximum teachers and students believed that the HSC test score did not reflect students' English language ability. This is quite strong evidence that the HSC examination in English exerts negative washback on teaching and learning English.

As a matter of fact, communicative testing is a challenge for test designers. One reason is the issue of predictive validity. When designing a test of communicative ability, identifying test takers' needs based on communicative encounters that they are likely to experience is one of the basic principles. However, it is not certain if test makers can guarantee that testees performing well on a test in class are also able to do well outside the classroom in a real life situation. One reason for this is that real life communication is characterized by unpredictability. Studies have proved that test designers have tried to make real-world tasks, but encountered difficulties from the varied or diverse nature of contexts (Katsumasa 1997; Brown, 2003).

5.3.4 Analysis of the HSC Answer Scripts in English

Answer script analysis is a part of portfolio assessment. It is a form of authentic assessment in which a student's progress is measured qualitatively. Answer script analysis can include evidence of specific skills and other items at one particular time and language performance and progress over time, under different conditions, in all four modalities (such as reading, writing, listening, and speaking) or all three communication modes (interpersonal, interpretive, and presentational). Using a combination of testing instruments lends validity and reliability to the portfolio.

Cheng (2004) suggests that analysis of answer sheets/scripts reflects students' overall achievement in second or foreign language learning. Like classroom observation, answer sheet analysis is of great value. Bailey (1999) comments that answer sheet analysis is closely linked to instruction, which has two educational benefits. First, linking assessment to instruction means that what is being measured has been taught. Second, it reveals any weaknesses in instructional

practices. Andrew (2004) suggests answer per analysis promotes positive student involvement. It is actively involved in and reflecting on their own learning. It offers the teacher and the student an in-depth knowledge of the student as a learner. This means that the teacher can individualize instruction for the student. Weak areas can be strengthened and areas of mastery built upon. Learners are involved in this process of tracking their learning and can take control of their learning.

Li (2009) suggests that answer paper focuses how much positive or negative washback dominates the classroom activities. Brown (2000) opines that answer papers are the visible evidence of learner's learning outcome. Enright (2004) suggests answer pages highlight how much communicative competence has been achieved opposed to how much it is tested. The analysis of the HSC answer scripts in English has observed that examinations contain little reference to the knowledge and skills that students need in their everyday life outside the classroom; and they tend to measure achievement at a low taxonomic level. It was found that the test did not correspond to the curriculum objectives. With regard to English language testing, the two major skills: listening and speaking are always avoided in the examinations at the higher secondary examinations.

Language should be tested communicatively, which has some principles. The *first principle* is to "start from somewhere." The study of Katsumasa (1997) and Wesche (1983) show an agreement with this statement by confirming that when designing the test, test makers should state carefully what they expect test takers to perform when they use the target language in a specific context, which means that test writers must know what they want to test. Bailey (1998) provides an example to illustrate this principle, which is that even though conveying and capturing meaning while maintaining accuracy are two important elements in communication, if the tester focuses on checking the test takers' ability to convey meaning, then accuracy will not be put into the scoring criteria. Therefore, it is unfair for students if test writers take into account other unstated objectives when grading tests, which also negatively affects the test's validity.

"Concentration on content" is the *second principle* in designing a communicative language test. The content here refers to not only topics but also tasks that will be implemented. The tasks should aim to be authentic and have clear reference in reality. These tasks match students' proficiency level and age.

The *third principle* of communicative test design is "bias for best." This means that test makers should bear in mind that they should create a test that can exploit test takers' performance at their best. The work of Brown (2003) also supports this principle in designing the test. According to Brown (2003), "biased for best" is "a term that goes little beyond how the student views the test to a degree of strategic involvement on the part of student and teacher in preparing for, setting up, and following the test itself" (p. 34). To illustrate this third principle, Bailey (1998) suggests that when she gives a test that requires students to do a dichotomy, she notices that the text might be higher than students' proficiency level; therefore, she reads the text three times and also encourages them to ask for new words involved in the text.

"Working on washback" is the *fourth principle* of communicative language testing. In order to obtain positive washback, test writers should create clear scoring criteria that would be provided to *both* teachers and test takers. Course objectives and test content are also put into consideration in the hope of promoting positive washback. The final aspect of communicative language testing is that of assessment. Communicative tests should be assessed qualitatively rather than quantitatively (Morrow 1991). However, Morrow (1991) argues that answers to tests are more than simply right or wrong, and that they should be assessed on the basis of how far towards an approximation of the native speaker's system they have moved. Tests should reveal the quality of the testee's language performance.

5.3.4.1 Answer Scripts Analysis Checklist

As pointed out in Chapter Four, the present researcher analysed answer scripts of the HSC English examination in English. The present researcher examined the answered scripts to examine, (i) what extent the communicative competence was tested, (b) how the students responded to questions, and (c) how tentative washback might generated from this test. For this purpose, an analysis checklist was applied (Appendix-3D). The checklist was adopted in accordance with the guidelines of Morrow (1991) and Brown (2003).

5.3.4.2 The Procedures of Answer Scripts Analysis

The analysis had two parts: observation of the checking procedures and analysis of the scripts (examined). The researcher observed the checking and scoring procedures while the examiners were examining the scripts. The questions of the checklist were constructed on the basis of Brown (2003), and Ferman (2004).The researcher analysed 20 answer scripts taken from both *First Paper* and *Second Paper* of English of HSC Public Examination held in 2010. Of the scripts, 10 were randomly selected from *First Paper* and other 10 were taken from *Second Paper.*

The present researcher observed 4 EFL examiners while scoring/examining the scripts. In this section, the analysis results are discussed and presented. The examiners were appointed by the Dhaka Board of Secondary and Higher Secondary Education. By profession, the examiners were EFL teachers at the HSC level. After the scoring was over, the researcher took the answer scripts under his disposal for analysis purpose. Each examiner checked 5 scripts. The researcher conducted the analyses in front of the respective examiners.

Two examiners were selected from English *First Paper* and other two were selected from English *Second Paper.* The answer scripts were anonymous and information of candidates was recoded in Optical Mark Recognition (OMR) format. The analysed answer scripts were taken from the examination held in 2010.   The present researcher observed how the examiners checked the scripts, how they marked each and every task and item, and how much time they spent for checking each script. The examined scripts were then analysed on the basis of checklist. The main points the researcher kept in mind that whether the answer scripts were the refection of the students' learning English, and whether students' performance was the ultimate outcome of washback of the HSC examination. The post analysis interview with the examiners was also conducted to collect their views on the examination systems contents of test, scoring, the students' performance, etc. The answer scripts analysed were coded (e.g. S1, S2, S3—S10), and the 4 examiners were given pseudonyms (e.g. E1, E2, E3, E4).

5.3.4.3 Guidelines for the Examiners

 In Bangladesh, the examiners are not provided any written criteria or guidelines to examine the scripts. An analysis of the HSC examination in English

subject indicated that its instructions were not clearly specified with respect to the procedures for scoring the test. No examples were provided as to how to answer a question. The examination paper included the credit limit at the right hand margin for each question. The examiners had no training in examining scripts. It was found that the education board organised a short (an hour) orientation program for all examiners on the day of distribution of the scripts. The orientation program mainly discussed administrative procedures such as date of submission of the examined scripts, date of publication of results, the remuneration of the examiners, and very little about scoring. There were no specific scoring guidelines or rules for the examiners. However, the interview results with the examiners revealed that the education board authorities instructed the examiners to evaluate the scripts liberally.

It was found that the HSC English examiners examined the answers as per their own evaluation ability, experience, and perception. Since there were no set guidelines, the evaluation and scoring of the answers differed from examiner to examiner. Therefore, the reliability of the score given to each answer was inconsistent. For example, questions in the Part-B (carrying 60 marks) of Second Paper, and the questions in the Part-C (carrying 40 marks) of *First Paper* were wrongly evaluated. There were possibilities of over-marking or under-marking to the answers. It was also found that the test score largely depended on the will, ethics, sincerity, and dedication of the examiners to their profession.

5.3.4.4 Reliability of Examining/Scoring of the Answer Scripts

The present researcher checked the reliability of the present scoring/examining system by checking and re-checking methods. He requested two head examiners of English First Paper Second Paper, to help the researcher check the reliability of the scoring of the answer scripts. The head examiner for English first paper called 3 examiners working under him and asked them to evaluate 2 answer scripts separately. The three examiners examined the same answer scripts (using a pencil). When an examiner completed the evaluation of the two scripts, it was given to another examiner to examine. The marking sign was erased before it was placed to another examiner. Thus, three examiners examined 2 scripts one after another in different time. The procedures of examining and re-examining of the scripts were not disclosed to the examiners. Here, the examiners were coded as: E1,

E2, and E3. The answer scripts were marked as script -1 and script -2. The scores given by the examiners are displayed in the table below (Table 5.75):

Table 5.75: Reliability of scoring - *English First Paper*

English Second Paper: Marks 100	E1	E2	E3	Mean	STDV
Script 1 (%)	78	65	70	71	6.55
Script 2 (%)	71	60	76	69	8.18

In the case of Script-1, the highest mark was given by E1 (78); the lowest score was given by E2 (65), whereas E3 gave 71 marks. The mean score of the script (Script-1) was 71 and standard deviation was 6.56. The difference between the highest score and the lowest score was (78-65) 13. On the other hand, in the case of Script- 2, the highest mark was given by E3 (76); the lowest score was given by E2 (60); whereas the mean score for the Script-2 was and 69 standard deviation was 8.19. The difference between the highest mark and lowest mark for the script (Script-2) was 16.

The scores for the first paper in both the scripts varied examiner to examiner to a large extent. The results were not consistent, and therefore, the scoring system was inconsistent with very poor reliability, which is a threat to the education system in Bangladesh. The table below (Table 5.76) shows the different scores given by the three examiners of English *Second Paper*:

Table 5.76 Reliability of scoring -*English Second Paper*

English Second Paper: Marks 100	E1	E2	E3	Mean	STDEV
Script 1 (%)	68	76	63	69	6.56
Script 2 (%)	63	69	74	68.66	5.50

The scoring reliability of English *Second Paper* was also tested in the same way as it was done for the first paper. It was found that the reliability of marking was seriously poor. The results were inconsistent because the score varied examiner to examiner. For the Script-1 (second paper), the highest score was given by E2 (76) while the lowest score was 63 given by E3. The mean score was 69 and STDV was 6.56. The difference between two scores was 13 marks. On the other hand, in the case of Script-2, the highest credit was given by E3 which was 74; the lowest score

was 63 given by E1. The difference was 11 marks. The mean score and STDV were 68.66 and 5.50 respectively.

It was a serious concern that, for the same answers, different examiners gave different marks; and the marks varied in high rate. While communicating one of the deputy controllers of examinations of Dhaka Board about the inconsistency of scoring, he commented that it was unfortunate to find such an inconsistency in the evaluation and marking procedures. However, he believed that, the marking might vary examiner to examiner in a tolerable level that must not over 5%. It meant that the results of the HSC examination were not reliable.

5.3.4.5 Findings of Answer Script Analysis

The researcher observed scoring procedures and conducted analyses of 20 answer scripts of 4 examiners (e.g.   2 examiners were observed for English first paper, other 2 examiners were observed for English second paper). While observing the marking of English *First Paper* scripts, it was found that the examiners spent only 4 to 6 (including filling the OMR form) minutes to examine and marking a 100-mark script. For the F*irst Paper*, the mean score was 63.4 and STDV was 14.73. It was a three-hour test; the examinees were to answer 14 questions (most of them had sub-items). For the benefit of the presentation of the finding, the answer scripts are coded as S1, S2...S10. Two examiners examined the following 10 scripts (from S1- S10).   The examiners were also given pseudonyms (for ethical reason) as E1 and E2. Examiner 1(E1) examined 5 scripts (S1- S5) while E2 examined other 5 scripts (S6-S10). The Table 5.77 displays the marks obtained and the amount of time spent for answer evaluation:

Table 5.77: Marks obtained and time analysis - English *First Paper*

English First Paper- Full Marks 100											
SL	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	Mean	STDV
Marks (%) Obtained	68	75	55	63	72	81	42	68	74	36	63.4	14.73
Time spent/ script(minutes)	4	5	4	4	6	6	3	5	6	5	4.8	1.03

As discussed in the earlier section, English *First Paper* comprises three parts: Part-A, Part-B, and Part-C. The study found that the examiners did not go

through the answers minutely and they gave very little attention to the answers. Most of the examiners looked at the length of the answers, underlined some of the words and sentences, and then put the credit at left-hand margin of the script. It was found that the scores of 10 scripts ranged from 36 to 81, whereas the averaged (mean) mark was 63.4. It was also found that the examiners took 3 minutes to 6 minutes (including filling up OMR sheet) to examine a script. On an average, the examiners spent 4.8 minutes to evaluate scripts. Figure 5.89 presents the findings of the analysis of answer scripts of *First Paper*:

Figure 5.89: Score and time analysis - English *First Paper*
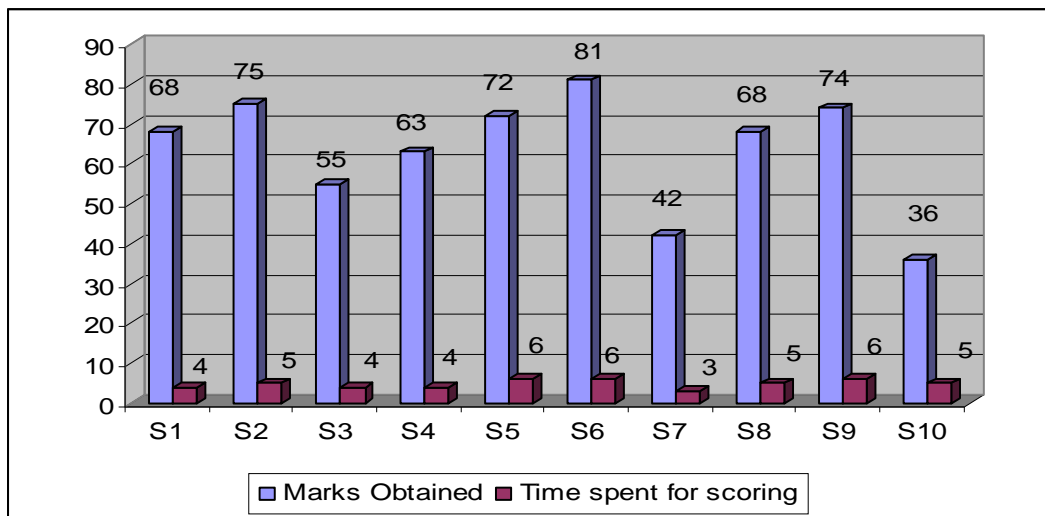
It was already found the evaluation procedures suffered from reliability, and scores given was in consistent. The examiners were found reluctant to examining answer scripts. They also seemed to be in a hurry to finish evaluation as early as possible. Since the examiners were reluctant to examining the scripts, since they did not have any training and guidelines for scoring the scripts, and since the evaluation/marking was found undependable/ inconsistent, the test scores of the HSC examination did not give the true picture of success.

In the figure above (Figure 5.89), the blue columns indicate the marks obtained and the magenta columns point out the time spent for scoring of an individual script. The researcher also observed 2 examiners during their examining of English *Second Paper*. When the examining was over, the researcher took the answer scripts for analysis purpose. The examiners gave their full consent to analyse

scripts in front of them. The Table 5.78 illustrates the students' scores and time
spent for marking each script:

Table 5.78: Marks obtained and time analysis - English *Second Paper*

English Second Paper- Full Marks 100												
SL	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	Mean	STDV
Marks (%) Obtained	69	73	53	67	72	51	80	77	61	65	66.8	9.57
Time spent/ script (minutes)	4	4	3	2	3	4	4	3	3	2	3.2	1.05

The Figure 5.90 depicts the individual score of the scripts along with the
time the examiner spent for checking and marking each answer script of S*econd
Paper* of English. The blue columns indicate marks obtained and the magenta
columns point out the time spent for marking each script. The examiner spent only
2- 4 minutes to examine a script:

Figure 5.90: Score and time analysis - English *Second Paper*
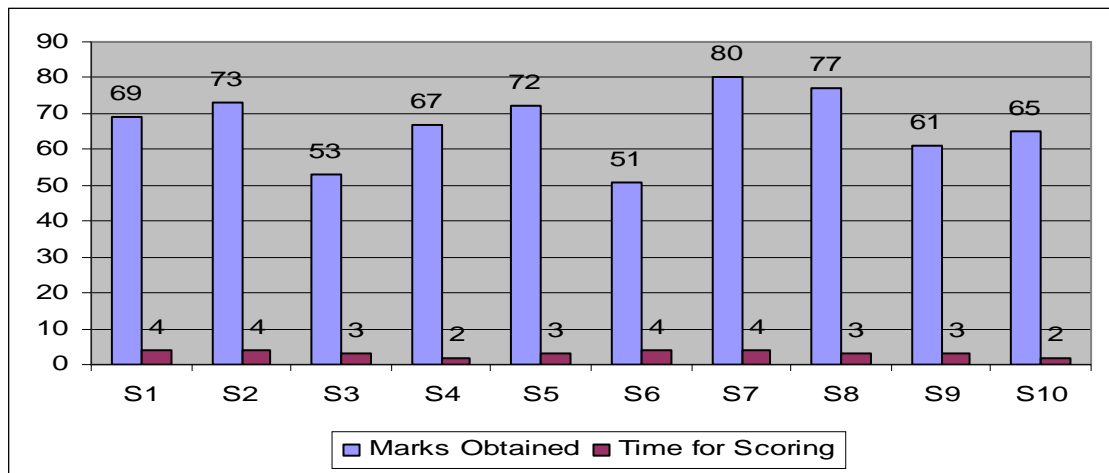
English *Second Paper* has two parts: Part –A (Grammar) and Part-B
(Composition). It was very frustrating to observe that the examiners evaluated the
five composition questions (Part-B) in less than half a minute (20 seconds- 40
seconds). The examiners just turned over the pages, underlined some sentences
without going through it, and completed marking. It was discouraging to observe the
checking and marking procedures of all the 4 examiners of both papers. For the
evaluation of answer scripts of English *Second Paper*, the examiners spent 3.2
minutes, on an average, of which half of the time they spent for filling up the OMR
sheet (also called top sheet). One of the examiner commented that the board

authority instructed them to check the scripts liberally. He also pointed out that there were lots of mistakes in the answers. "If I check it properly it will take long time and most of candidates will fail", he added.

5.3.4.6 Skills and Linguistic Elements Tested

The analysis of the scripts found that the grammar, vocabulary, matching, completing sentences, and composition writing were tested. They were all about language elements and writing skills. The analysis found there was no unified system to answer the questions, for example, in the case of *Second Paper* (questions 1-8), many students only wrote the missing words or correct words/form of verbs, whereas some students wrote full sentences. The examiners gave equal credit for all styles. It was found that there was no provision of testing the listening and speaking skills in the present examination system. However, reading comprehension was tested in the examination, for example, in the first paper, the questions from 1-8 were reading comprehension related. The analysis observed that the students got 60%-80% marks in this section (Table-5.79).

5.3.4.7 Maximally Attempted Questions

The analysis of the answer scripts found that 90% students answered all the questions. The grammar and short items were the most favoured tasks for the students. More than 50% students did not maintain the sequence of the question number. There is no obligation to maintain the question serials.   In the second paper, 10% students avoided the direct and indirect speech (Question No. 5) item.

5.3.4.8 Items Attended First

All the examinees started their examination with answering the grammar, vocabulary, or objective questions items. The reason was that the grammar and language elements were usually and repeatedly practised in the class during the whole two years of time. Those questions were also considered easy scoring items

5.3.4.9 High Scoring Questions

For English *First Paper*, the analysis of the answer scripts found that grammar, vocabulary and multiple choice questions brought maximum marks. The table below (Table 5.79) shows the areas where the students did the best. In Part – A, on an average, the students got 28.5 marks ranged from 20 to 35 (out of 40):

Table 5.79: Marks obtained in the different parts – English *First Paper*

Parts/ Areas	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	Mean	STDV
Part -A=40 **Seen Comprehension** (Multiple choice, true/ false, Filling the blanks, etc)	30	32	22	25	32	35	20	28	31	30	28.5	4.76
Part -B=20 **Vocabulary**	**20**	19	15	16	**20**	**20**	12	**20**	18	15	17.5	2.83
Part -C=40 **Guided Writing** (Rearrange, matching constructing paragraph)	18	24	18	22	20	26	10	20	25	29	21.2	5.32
Total	68	75	55	63	72	81	42	68	74	74	Average	

The part B received the mean score 17.5, and standard deviation was 2.83. In this section, the highest marks obtained was 20 (out of 20), and the lowest score was 15 marks. The vocabulary section (Part-B) got highest score, while seen comprehension section received the second highest position. The findings of the classroom observations, questionnaire surveys, and interviews revealed that the most practised items in the classroom were vocabulary, multiple choices, true/false and grammar.

These items are repeatedly practised in the classroom, which is the observable evidence of washback of the HSC examination on teaching and learning. Washback is not necessarily unidirectional, i.e. from exam to textbook and teaching rather than bi-directional, i.e. also from textbook and teaching to exam (Wall 2005, and Hawkey 2009). In Figure 5.91, the part-wise marks of 10 scripts are displayed:

338

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

Figure 5.91: Marks obtained in section - English *First Paper*
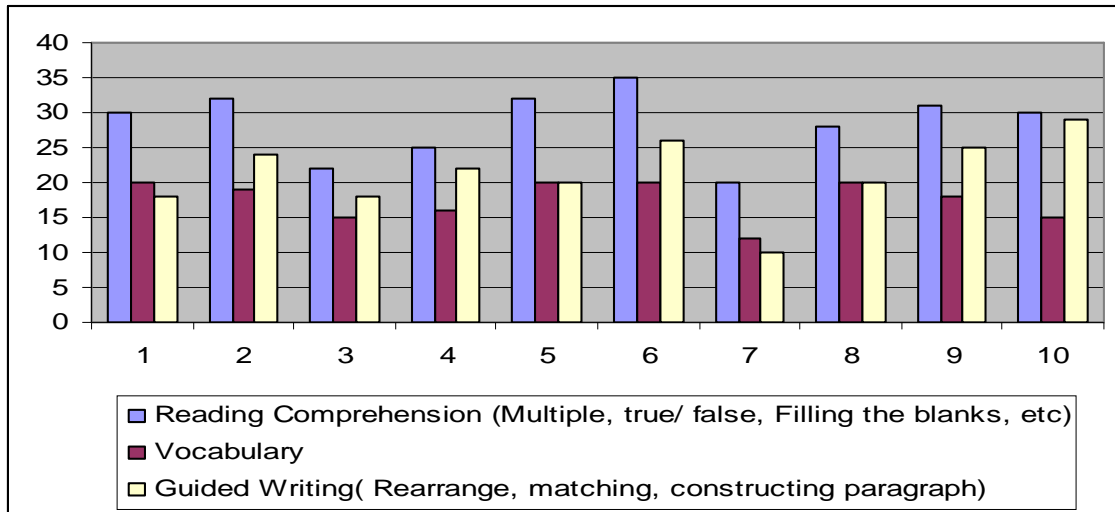
The blue columns indicate seen comprehension; the magenta columns show vocabulary marks; and the yellow columns present guided writing marks. The first script (S1) received highest marks in the reading comprehension section among all three sections. Seen comprehension carried 60 marks. Therefore, this section was the highest scorer in all 10 scripts, but the vocabulary section received the highest average (17.5 out of 20 marks). The figure (Figure 5.91) shows, in the vocabulary section, 4 scripts got 20 out of 20 which is 100% success.

In the case of English *Second Paper*, the grammar section has relatively stronger mean. On average, students got 33.5 marks out of 40, which indicates that students performed relatively well in this section. The analysis found that the students got 33.3 on an average (out of 60) in composition section. In this section, marks varied from 23-42. This variance is relatively very high. The table below (Table 5.80) presents the findings of section score analysis:

Table 5.80: Marks obtained in the different parts -English *Second Paper*

2nd Paper	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	Mean	STDEV
Grammar Marks 40	34	38	30	32	35	28	38	40	28	32	33.5	4.25
Composition Marks 60	35	35	23	35	37	23	42	37	33	33	33.3	6.00
Total	69	73	53	67	72	51	80	77	61	65		

Figure 5.92: Marks obtained in section- English *Second Paper*
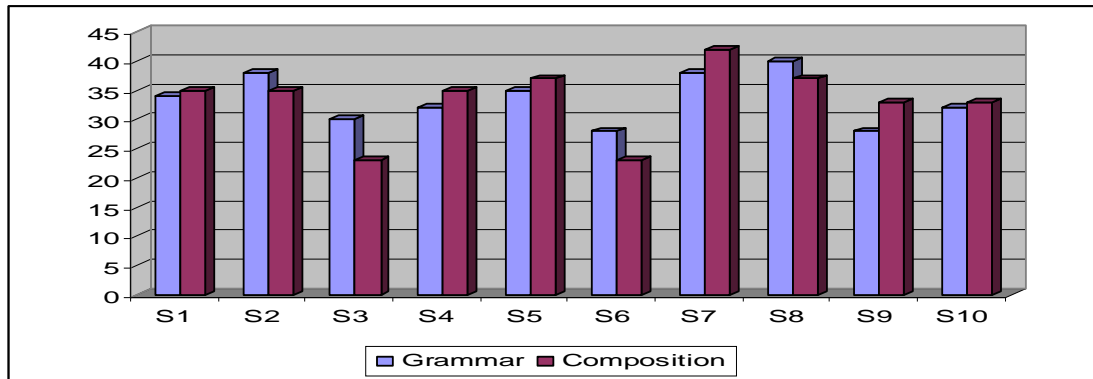
The Figure 5.92 projects how the students performed in the different sections of the examination papers (question paper). The blue columns indicate marks obtained in the grammar section, whereas the magenta columns show marks obtained in the composition section.

5.3.4.10 Evidence of Washback from Answer Scripts Analysis

Portfolio such as the answer sheet is a true evidence of washback (Ross, 2005).   Analysis of test scoring is a strong means of investigating washback on language instructions and learning (Hawkey, 2004). The present researcher analysed 20 answer scripts of the HSC examination in English. The study revealed some truth. It was found that only reading comprehension, writing, grammar and vocabulary are mainly tested in the examination, though the curriculum has emphasized the communicative competence. The textbook reflects the goals of the syllabus and curriculum, but the teachers were found reluctant to teach communicative English; it was because the communicative language was not at all tested in the HSC examination.

The examination system is overall responsible for generating negative washback on ESL/EFL education (Shohamy, 2005). The answer scripts analysis found that the examiners hardly cared about the proper evaluation of the scripts; they simply performed the duty of giving the marks for answers. They spent only 2-5 minutes to examine a 100- mark script. The study disclosed the fact that the students preferred to answer the grammar and vocabulary items because they could achieve relatively higher score in those items. It was found vocabulary, grammar

340

items, true/ false, matching, rearranging all were selected and taken from the prescribed book, *English for Today* (for classes 11-12). Over-marking was found a common practice to all the examiners.

The examiners disclosed that they were instructed by the education board authorities to examine the scripts very liberally. It indicated that the authority wanted to increase to the pass rate quantitatively. In addition, most of the cases, the questions were over-weighted with excessive marks. For example, in the first paper, a question of matching phrases of 6 sentences carried 12 marks; a question of just rearranging of 14 sentences in a proper order carried 14 marks; a paragraph (answering five simple questions) in 150 words carried 14 marks. All these questions were extracted mostly from *English for Today* and directly commercially produced test related materials such as guidebook, suggestion book, test papers, etc. The test contents, examining/scoring procedures, marks allocation to the question, poor test constructs, etc. generate negative washback on teaching and learning at the HSC level.

Many researchers have claimed that high-stakes testing might trigger a myriad of unethical test preparation practices or motivate teachers to manipulate students' test scores (e.g. Ferman, 2004; Hawkey, 2006). To date, researchers have paid most of their attention to the washback of tests on four domains of teaching practice:(1) content of teaching, (2) teaching methods, (3) assessment methods, and more broadly (4) overall teaching style, classroom atmosphere and teachers' feelings towards the test. Among the four domains, teaching content was always found to be altered by tests.

Due to poor reliability of test cores, the pass rate of examination cannot draw a true picture of HSC EFL education. In a particular case, while writing a paragraph of 150 words, a testee committed 21 spelling mistakes, and produced all (total 9 sentences) linguistically incorrect sentences. An examiner gave 80% marks for writing this paragraph. The researcher was informed all the examiners were EFL teachers in different higher secondary colleges. In the language classrooms, the teachers practised selected items again and again where there was less risk to score higher. It can then be concluded that the present HSC examination in English generates negative washback on teaching and learning English as a foreign language (EFL).

5.4   Findings of the In-depth Interviews

This section reports and discusses the findings of the interviews with EFL teachers, EFL examiners, and curriculum specialists. This was an interview on a one-to-one basis. It was designed to elicit qualitative data on: how they planned, how they designed the policy, how they delivered inputs, and how they received outcomes. In order to triangulate and possibly extend the findings of the present study, the present researcher conducted semi-structured interviews with 6 EFL teachers, and 4 HSC examiners of EFL, and 3 curriculum specialists. They were all directly involved in HSC education in Bangladesh. First, the findings of interviews with the EFL teachers are presented. Then, the researcher records the results derived from interviews with the EFL examiners, and finally presents the interview results with the policy makers. The presentation of results is followed by discussions.

5.4.1   Interviews with the EFL Teachers

The role of teachers as one of the main stakeholders in the implementation of educational change and curriculum innovation has been the focus of ongoing interest to curriculum researchers, and has been examined extensively in the literature, both conceptually and empirically. This is because teachers, as implementers, determine whether or not curriculum innovation is executed in the classroom as it is intended by policymakers (Carless, 1999). For instance, Carless (1999) asserts that teachers are the individuals who implement, adapt, reject, or ignore curriculum innovation. The objectives of the syllabus and curriculum are gained through their classroom activities. They are main agents to generate positive and/or negative washback from test on teaching and learning.

Among the factors that can mediate the washback effect is the teacher (Wall, 1996) and her/his perceptions about the examination, its nature, purposes, relevance in the context, etc. Teacher perception, teacher attitudes and teacher beliefs are often mentioned in the washback studies as well (Rea-Dickens, 1997; Cheng, 2004; Watanabe, 2004). Therefore, teachers' decisive roles in the implementation of the curriculum cannot be underestimated. Without their support and genuine involvement in the innovation, any curriculum implementation will stay at a superficial level, with either semi- or even non-implementation. This section

discusses the disjunction between policymakers and implementers as well as implications for teachers, based on the existing research data.

The findings of the previous researches on teaching show that washback is subject to how (methodology) teachers teach. This may be attributed to Hawkey's claim (2006) that "the distinction between course content and methodology is not always clear cut" (p. 106). The examination had had considerable impact on the content of English lessons and on the way teachers designed their classroom tests (some of this was positive and some negative), but it had had little or no impact on the methodology they used in the classroom or on the way they marked their pupils' test performance. Therefore, it was very crucial for the present study to conduct interviews with EFL teachers to examine whether their perception of test, classroom behaviours, curriculum knowledge, etc contribute to generating washback on teaching and learning.

All the teachers interviewed were the participants of classroom observation. During the classroom observation the present researcher observed 10 EFL classes of 10 English teachers. Of them, the researcher selected 6 EFL teachers for interviews. Of the teachers interviewed 3 were male and 3 female. The teachers were coded as T1, T2 ... T6. The profiles of teachers were stated in section 5.2.2. The teacher interviews were based on the classroom observation they participated in. Therefore, the results of teacher interviews corresponded to the findings of classroom observation (Chapter 5.2). Semi- structured questions (Appendix- 4A) were used for teacher interviews.

5.4.1.1 Interview with Teacher 1 (T1)

The researcher conducted interview with T1. He along with his students was observed during classroom observation. He also took part in the questionnaire survey. Two days before the classroom observation started, T1 initially impressed the researcher as an advocate of student-directed instruction. He explicitly mentioned that the way he taught was in line with his teaching beliefs and the needs of his students. The present researcher had expected that he would incorporate some learner-centered activities in his class. Contrary to expectations, no such activities

were observed throughout the observation process. In fact, his class turned out to be typically teacher-centered.

In the class, he was observed spending a lot of time lecturing on linguistic knowledge. Pair work/group work activities were hardly implemented. Meanwhile, although the researcher observed him occasionally talking vigorously to his students by asking them comprehension questions related to the texts being taught, his instruction was conducted in Bengali. Moreover, he supplied them with answers most of the time. The results reveal a big mismatch between what he said and what he practiced. One possible reason for such a mismatch as well as his extensive use of Bengali in the classroom was that his own English language proficiency level was not very high, which made it impossible for him to adequately express his ideas in English. The researcher made such an inference based on the statements he made at one individual interview.

When addressing the role of language proficiency in EFL instruction, he stated that he occasionally found it hard to convey his ideas in English. Another possible reason was that although he previously expressed his interest in CLT and task-based teaching approaches, he might not be well-equipped with the hands-on knowledge which would allow him to manipulate the approaches as he wished. Evidence for this inference can be found in the conflicting remarks made by him on different interview occasions. During the interview, he was talking in the presence of his colleagues, he criticized the structural approach (or grammar-translation approach) saying:

> This approach is too stereotyped and backward. I believe that an English class should revolve around productive skills, since the goal of our teaching is to enhance students' communicative competence.

In replying to a question, he articulated some interesting beliefs in teaching communicatively:

> Although helping students acquire communicative competence is our goal, it seems unattainable. Therefore, pure communication is deceiving. Above all, students need to lay a solid foundation in grammar.

5.4.1.2 Interview with Teacher 2 (T2)

T2 was interviewed 2 days after she had taken part in the classroom observation. It was found that T2 taught her lessons using the structural approach. The instructional pattern she followed was a monotonous one, starting with reading the text aloud herself, then presenting and explaining language points (e.g., vocabulary and grammar), and ending up analysing the text paragraph by paragraph and translating them into Bengali. One strong impression from the observations was that the teacher's lecture dominated the class. The researcher observed the students being ignored most of the time and rarely called upon in class. Interestingly, T2 herself admitted that her classes were very much teacher-centered. When asked why no group-work activities were organized, she said:

> I have tried using group work, but I find that such kinds of activities are ineffective. It is a waste of time to conduct them. The reason is that instead of using English, my students tended to talk to one another in Bengali.

She also expressed her dilemma as to what to teach in the class by saying:

> I used to teach by focusing on the meaning of the text rather than on language points, but my students complained about the meaning-based instruction saying that they had not learned anything. As a result I had to concentrate more of my class time on teaching vocabulary and grammar. I sometimes feel I do not know what to teach and how to teach.

Her statements seemed to imply that whether or not teachers could organise communicative activities depends on the motivation as well as the English level of their students. She also seemed to suggest that students' beliefs or students' roles in the classroom constitute a barrier to the implementation of communicative activities in their instruction. This view was corroborated by T1 who also pointed out that students preferred to be taught more vocabulary because it is tested in different forms.

In addition, like T1, she reflected on the challenge she confronted when it came to organising communicative activities. From her accounts, her own low oral proficiency also poses a constraint on her instruction. She also echoed T1's claim saying that she sometimes found it hard to convey her ideas freely in English as well. When accounting for the reason that led to teachers' low oral proficiency, she articulated:

This has a lot to do with the policies and orientations of the colleges as well as the government. As a rule, little attention has been given to whether you teach well or not. As long as you have published a certain amount of journal articles and done well in research, you are considered having accomplished your job.

She further reminded the researcher that similar challenges were also faced by many other EFL teachers like her.

5.4.1.3 Interview with Teacher 3 (T3)

The researcher interviewed T3 at his college campus. T3 wanted to talk before his colleagues. He claimed that he was a very serious type of teacher and most of his students did well in the examination. It was found that T3's lessons were also characterized as being knowledge-oriented and teacher-dominated. Reading is the primary skill emphasized by him. The focus of his instruction was predominantly on language knowledge. Teacher talk took up 60-70% of his class time. What struck the researcher was that as part of the class routine, he invariably started each new lesson leading his students to go over the vocabulary lists provided at the end of each text (lesson) before giving his lecture on the text. Apart from activities such as reading texts aloud and translation, rarely he was observed interacting with his students for the purposes of communication. Throughout the observation process, he talked about the importance of students' participation in classroom activities, but pair-work or group-work was never observed in his class. Furthermore, she seldom produced extended sentences in English. Much of the classroom instruction was carried out in Bengali.

He organized test-related activities, and the method he used was stereotyped. For example, when he led his students to do fast-reading exercises, first he simply asked the students to quietly read a passage he passed out to them and then answer the given questions. After that, he checked the answers with the whole class and provided them with the keys by highlighting the essential vocabulary and explaining why each choice was made. Between whiles, he gave the students tips on how to deal with the similar types of questions in case they appear on the HSC examination, which was the direct evidence of negative washback of examination.

When the researcher asked T3 to comment retrospectively on the teaching strategies he utilized in his instruction, he articulated some interesting beliefs in why he taught vocabulary this way. Here is an excerpt from his remarks at the interview:

> Seldom did I think of such issues as teaching methods. I simply taught using my own way of teaching. Personally I think that teaching priority should be given to developing students' abilities in reading, because I find the biggest barrier the students encountered while reading is that their vocabulary is limited. Consequently, they had difficulty understanding the passages they read, and furthermore they could hardly convey their ideas when writing compositions. Therefore, linguistic knowledge should still be stressed.

He further justified his practice by saying:

> The students are used to the method of structural analysis, and they find it hard to change their traditional way of learning. Each text of *English for Today* for classes 11-12 consists of a large number of new words. If we do not explain them, the students do not know how to use them.

He also defended his use of Bengali in the class saying, "Using English is a waste of time. The students may not be used to being taught purely in English in class". The questionnaire surveys and classroom observations also revealed that the English teachers used Bengali as a medium of instruction. One of the reasons of using Bengali frequently in the classes was that they themselves were not adequately fluent in English.

5.4.1.4 Interview with Teacher 4 (T4)

The researcher interviewed T4 in the teacher's room of her college. She showed her sincerity and expressed her willingness to cooperate the researcher in providing information she knew. During classroom observation, she was found very lively and friendly in her class. Unlike other participants who attached more importance to language forms, T4 stressed the development of students' ability to use English. She was so highly motivated that she spontaneously experimented with communicative activities as well as cooperative learning activities (e.g., pair work/ group work, language games, questions and answers) in her classes. Not only was she observed frequently utilizing authentic materials, but she was also found using textbooks more creatively and trying hard to encourage her students to interact in class. It was noted during the observations that her students showed higher

motivation in learning English and were more active in class than those of other classes observed.

When recounting the reasons why she implemented these interactive activities rather than spending a lot of time on language forms, she articulated:

> I attempted to arrange as many activities as possible, because I was afraid that my students would be bored with my lessons. What I cannot stand is that they all lower their heads and do not respond to me. I do not think the "Cramming-Duck" method works. I believe that if a teacher lectures for two hours, a student will only end up acquiring 10% of what he or she is taught. In my opinion, teaching vocabulary as discrete points does more harm than good to the students, even though they prefer to be taught that way. The more vocabulary we explain, the more confused the students will become. It is impossible for students to have a command of it by learning it in such a decontextualized way.

The above comments provided insight into her professional stances on EFL instruction. In replying to a question, she pointed out that she was aware that most of her colleagues continued to devote plenty of their class time to teaching language forms. She proceeded to pinpoint the reasons why teachers at large had a preference for teaching vocabulary items. The following is an excerpt from her interview discussion:

> There are a number of reasons for this. One is that they may have been taught that way. Another plausible reason is that they are constrained by their own language proficiency. To be specific, they have trouble expressing their ideas in English themselves. In such a case, teaching vocabulary is the easiest way they can do. By so doing, they no longer need to take the trouble to improve their own language abilities. Nor do they need to rack their brains on how to teach. Another reason might be that that way of teaching, in their opinion, appeals to their students.

She commented that the majority of teachers had never thought of whether it was appropriate to teach vocabulary that way; they simply followed what other teachers did. She added that it was also possible that some teachers were not responsible or conscientious enough. In spite of her efforts, her class still seemed deficient in that she was seldom seen calling on students to answer her questions on a one-to-one basis. She was also aware that she had encountered some obstacles while carrying out student-centered activities. With respect to the impact of learner beliefs in teaching, her opinions are consistent with those of others. One example

she gave the present researcher corroborated T1's and T2's assertion that students' beliefs also had a part to play in the way that teachers taught:

> Once I received a letter from my students saying that they enjoyed my class very much. While they assured me that they were contented with the way I delivered my lessons, they expressed their worry about the group discussions that I had assigned them to prepare, for they reminded me that their exam was round the corner.

This example serves to illustrate that students' beliefs also play an important part in teachers' decisions as to how to teach. In any way, students sometimes influence the teachers to teach particular items important for the examination which indicates the unavoidable washback of HSC examination in English on teaching and learning English.

5.4.1.5 Interview with Teacher 5 (T5)

As mentioned earlier, the interviews with the teachers were based on the classroom observation. So, the findings of interview cannot be isolated from the findings of classroom observation. The observation data showed that the way T5 dealt with her lessons exhibits features of both traditional methods of teaching and CLT. She was observed using her textbook creatively by going beyond it to create local contexts for her students to use the language. At such times, she was carrying out activities to practice students' skill in speaking. She was found using the textbook in a formal way, dealing with it as a means of reinforcing language forms such as vocabulary and grammatical points. During these times, a lot of translations and paraphrases were utilized. When asked to explain her reasons for using methods of translation and paraphrases, she told:

> I frequently put what I said from English into Bengali to highlight what was taught. In this way, I can clarify what the students may not have understood. With reference to paraphrasing, It is hard to say whether it is good or not. In many cases, teachers have prewritten the paraphrased sentences on their own textbooks. Generally the sentences are copied from teachers' books. In class, they simply need to read them.

She reflected retrospectively on her own way of teaching, saying:

> Although I prefer to use CLT and attempt to conduct meaning-based instruction, the time I dedicated to preparing my lessons was, to be frank, quite limited. Had I committed more time and energy to my teaching, my lessons would have taken a dramatic turn.

During the observations, she was found devoting a lot of time to explanations of linguistic points. She defended why she was doing so, saying:

> Language is a means of communication. When students talk, they need to demonstrate a good command of linguistic knowledge. Otherwise they will be at a loss what to talk about. In my view, language use should take place under the condition that there is some real content. Content is the carrier of communication.

5.4.1.6 Interview with Teacher 6 (T6)

Before the classroom observation started, the researcher informed the teachers that he would talk to them to share views. Accordingly, the researcher interviewed T6 on his classroom teaching. T6's teaching patterns could be characterised as combining aspects of the traditional method (e.g., with a focus on basic skills such as pronunciation and recitation) and CLT (e.g., engaging students in discussions and negotiation of meaning).

During the observations, he was often seen asking students to read texts aloud and in the meantime modeling correct pronunciation. While he placed special emphasis on pronunciation, the time he spent on teaching vocabulary and grammar was much less than that other observed teachers did. Compared to other observed teachers, he used more English in class. Similar to the students of T4's class, his students were all well-disciplined and highly cooperative in the classroom, which might be related to the high expectations he set for them. A look at the interview data demonstrated that his practices in the classroom reflected his teaching philosophies. The following comments reflected some of his teaching stances:

> To acquire a language is to use it. So, our teaching should aim at helping our students to acquire the competence to use English. Foremost, we must help them lay a solid foundation in basic skills such as pronunciation and talking in appropriate English. To develop these skills, memorisation is pivotal. Without memorizing some standardized texts or articles, it will be impossible for them to express their opinions freely. That is why I assigned my students 42 topics and each class each student is supposed to be able to recite 3 paragraphs in relation to these topics. My students are cooperative, because I let them know my purpose of doing so. It is important to communicate with students and let them know how to learn.

It can be seen that based on his notion, the ability of language use can be acquired through practice of basic skills such as pronunciation and rote memory. An

interesting finding is that like T4, T6 was negative about devoting a lot of time to teaching vocabulary items, saying that it was quite time-consuming and impeded the enhancement of students' language competence. According to him, even if the students were taught a lot of vocabulary, they could hardly remember them. Like T4, he made an interesting comment on why many teachers prefer to teach vocabulary:

> Quite a number of students believe that the teachers who lecture on vocabulary are both high-leveled and knowledgeable. But the fact is that teaching vocabulary is the easiest approach.

He also explained why he spoke English most of the time in his classroom. His assumption is that because he teaches in English, his students would be able to be exposed to the language as much as possible. In an individual interview, he revisited the issue of teacher quality touched upon by his colleague at the group interview. Some interesting points he made on the issue are as follows:

> EFL teachers' input in the classroom plays a crucial role in students' exposure to the language. It's more important than the exposure they receive when listening to the recordings or watching TV, since the interactions between a teacher and his/her students are more direct. If the teachers' oral proficiency has improved, then EFL teaching will be upgraded to a higher level. However, the reforms have been made to English second paper cannot attain such a goal.

The purpose of EFL teacher interview was to explore how the interviewed six teachers conducted their teaching with regard to the intended curriculum. From the interviews, the four EFL teachers (out of six) recognized their own lack of knowledge and understanding of the syllabus. Although the curriculum designers/ policymakers expected teachers to adhere to the objectives and specifications of the syllabus in their classroom teaching and to be knowledgeable and clear about the syllabus, the teachers also expressed their lack of interest in this curricular document. In terms of the student-centred approach, all the teachers interviewed attributed large class sizes, students' poor language proficiency, limited teaching periods, heavy teaching loads, and students' study habits as obstacles to their implementation of this approach. Maximum teachers stated that one hundred percent use of English in instruction would probably result in students' frustration, based on their students' language ability. They remarked that use of their first language was beneficial for their students.

Regarding the use of the mother tongue in the classroom, these four teachers emphasized that they used Bengali to save time, to be clear in instruction, and to check whether students understood what was being taught to them. The teachers acknowledged the impact of this test on their curriculum implementation. The effects included teaching to the test, a narrowing of the intended curriculum by focusing on improving students' test scores, and paying scant attention to the cultivation of students' communicative skills. Indeed, both classroom observations and interviews demonstrated a discrepancy between what was intended by the policymakers and what was enacted by the two teachers.

5.4.2 Interviews with EFL Examiners

The researcher interviewed 4 EFL examiners to obtain qualitative data on their perception, beliefs, and practice as examiners of English. Of the four interviewed examiners, two examiners evaluated English *First Paper* and the other two examined English *Second Paper*. Earlier, they were observed while examining and scoring the answer scripts. Observation took place at their respective colleges. After they had finished their evaluation, the researcher took the examined answer scripts for analysis. The examiners were requested to share views on their evaluation and scoring systems. The semi-structured questions (Appendix- 4B) were prepared earlier so that necessary information could be elicited.

The major purposes of the interview were to- (i) investigate whether any internal (their belief, perception. experience, knowledge, etc) and/or external pressures or factors influenced them in the process of examining and scoring the scripts, (ii) examine whether the answer scripts evaluation procedures/activities influenced their teaching and learning, and (iii) look into whether the present answer scripts evaluation systems contributed to generate washback (positive or negative) on classroom teaching and learning. The researcher interviewed the examiners on separate days and behaved as if the researcher were just sharing views with them. All the examiners had been English examiners for several years under the Board of Intermediate and Secondary Education, Dhaka. The interview was based on the observation of scoring procedures and analysis of HSC answer scripts in English. The findings and discussion of the interview are presented below. The examiners are coded here E1, E2, E3, and E4.

5.4.2.1 Interview with Examiner 1 (E1)

E1 was an examiner of English *First Paper*. The introductory question to him was - whether he enjoyed checking of the HSC EFL answer- scripts. E1 replied that he did not enjoy. He added that by examining scripts he could best use his idle time when there were no classes. He continued:

> It increases my acceptance to my students; it also brings some financial benefit; I think all the English teachers should examine board scripts because it helps how to prepare the students for the examination; and it helps me compare how the students in different colleges answer the questions.

When asked, whether he received any training in evaluation of scripts, he replied that he did not because there was no opportunity for training for the examiners to examine and mark the scripts. The researcher asked if he received any written guidelines for scoring, he said there were not such guidelines provided by the education board, but he took part in a short orientation on the day of distribution of the scripts. He said that the board authority instructed them to be liberal while examining the scripts, although how much liberal an examiner would be was not specified. When he was asked to comment on the question paper and answer script evaluation (especially on taking so short time for checking), he said:

> Most of the answers are of multiple choice types such as matching, vocabulary, and rearranging. So, it took very short time to evaluate a script. He continued saying that the questions 11, 12, and 13 were over weighted with excessive marks. He remarked that those items should be changed. The education board has instructed them to give credit for every question if the candidate can organise and generate an idea. 80% candidates can not write even a single correct sentence on a given paragraph. But he has to give them over 70% marks as per the instruction of the board authority.

In reply to another question, whether the present evaluation system was appropriate, he replied that the system was not at all authentic and appropriate. He further continued that the most important thing was that the test contents should be changed; the problems were with the question setting. He remarked, "This can yield high pass rate and good grades, but this system would eventually impede English education in Bangladesh".

From the interview with E1, it was found that the answer scripts assessment was not an interesting one; he examined the scripts for a number of reasons as he mentioned. He also expressed his dissatisfaction with the question paper and poor

performance of the students. The examiner suggested changing the contents of the HSC examination. From this interview, it was found that the present HSC examination system was inappropriate to enhance EFL teaching and learning. Andrews et al. (2000) point out the any inappropriate language test must exert negative washback on teaching. Barnes (2010) believes any language test should aim at promoting language learning. It is believed that test should aid learning, and a language test should be used as lever. Positive washback takes place when tests induce teachers to cover their subjects more thoroughly, making them complete their syllabi within the prescribed time limits

5.4.2.2 Interview with Examiner 2 (E2)

She was an examiner of English *First Paper*. She had been an examiner for last 10 years. Last year, she discharged duties as a head examiner. She informed that she enjoyed examining and marking the scripts. She added that the authority should have arranged some training for the examiners. When she was asked to comment on the present HSC examination, EFL education, and validity and reliability of answer scripts evaluation and marking. She commented:

> We always speak about CLT, but what the real picture is! English is neither taught communicatively nor is communicative English is tested. This is the examination system which is overall responsible for poor performance of the students in English in real life situation. Examination system is very poor; the questions are not maintained difficulty level; we cannot expect testing of students' EFL proficiency by introducing matching, rearranging, and true/false items. She posed a question- Do the policy makers really understand what CLT means and how it is tested? The pass rate is high; therefore, the government is highly pleased and overwhelmed.

In Bangladesh, all the EFL examiners are the teachers by profession at different colleges. It is a fact that an examination influences teaching and learning; additionally, when the authority asks the examiners to be very liberal while evaluating the answers, the teaching and learning further deteriorate. She said that she took 4-5 minutes to evaluate a script because the answers were very redundant and short. She believed that examiners should be provided guidelines for checking and marking the answers. She also believed that the scripts were not properly evaluated. She complained the answers were over-marked by the examiners. E2 believed that the present EFL examination system is inappropriate for language

testing. She further said, in the present examination system, especially in English first paper, no skills were tested except some language elements. She emphasised that the examination system should be revised introducing new items in the form of communicative language testing. She also advised that the government should at least study and look into the other South Asian countries' (e.g. India, Pakistan) English language teaching and testing system.

From the interview results, it was again found that the present EFL test was not a valid test as it failed to assess students' communicative ability in English. The findings of the interview with E2 revealed that the answer scripts were not evaluated properly. Good tests can be utilized and designed as beneficial teaching-learning activities so as to encourage a positive teaching-learning process (Pearson, 1988). A creative and innovative test can quite advantageously result in a syllabus alteration or a new syllabus (Davies, 1990). Cheng (2005) points out that an examination achieves the goals of teaching and learning, such as the introduction of new textbooks and new curricula.

5.4.2.3 Interview with Examiner 3 (E3)

E3 was an examiner of English *Second Paper* under the Board of Intermediate and Secondary Education, Dhaka. She had experience of examining the scripts of 4 years. She was asked if she enjoyed the answer scripts evaluation, she replied that in the affirmative saying that she preferred English *Second Paper* because it took less time than English *First Paper.* Like E1 and E2 she also informed that there was no scope for training for examiners. She admitted that the education board authority instructed them to be liberal while marking. The researcher observed her checking and marking sitting in front of her and found that she (E3) examined a script of 100 marks in 2-3 minutes. The researcher asked how she could finish checking in little time. She replied that she looked at the particular point of grammar items to check. She added that in the comprehension section the students made huge mistakes so she underlined maximum sentences. She pointed out that the candidates committed spelling mistakes, grammatical mistakes, etc. She continued by saying that:

> The present examination system is faulty. There is no sign of communicative testing. The question items are so funny. Questions are set to make the pass

rate high. The answer script evaluation is manipulated. The students make so many mistakes that they should not be given any credit on composition items; reading their answers line by line is simply waste of time. The answers are not suitable enough to get any credit. I look at how much the candidate has written of his/her idea. The education board authority instructed them to give credit for such a poor answer.

She made some more frustrating comments on examination while answering a question. She said that the examination system and the answer scripts evaluation procedures influenced her classroom teaching. She remarked that since she knew that there was sufficient flexibility in checking and marking the answers in the final examination, she remained relaxed to some extent and did not put serious attention to the accuracy of the students. The researcher asked whether the present examination system and its question items enhance EFL teaching and learning. In reply to this question, she said the present EFL examination system is an instrument to hamper language education. However, she emphasized that the examination system should be modernized to evaluate students' communicative competence. It was found that the examiners evaluated an answer script in 2- 4 minutes. The researcher found that the examiners just turned the pages and underlined some sentences and words, and then put marks on the left hand margin.

The findings of interviews disclosed the fact that the reliability of scoring was inconsistent, and the results from it could not give a true picture of EFL learning and language ability. Cohen (1994) describes washback in terms of" how assessment instruments affect educational practices and beliefs" (p. 41). Pierce (1992) specifies classroom pedagogy, curriculum development, and educational policy as the areas where washback has an effect. On the other hand, Alderson and Hamp-Lyons (1996) took a view of washback which concentrated more on the effect of the test on teaching. Positive washback would result when the testing procedure reflects the skills and abilities that are taught in the course, as, for instance, with the use of an oral interview for a final examination in a course in conversational language use. Test contents can have a very direct washback effect upon teaching curricula.

5.4.2.4 Interview with Examiner 4 (E4)

Examiner 4(E4) was an examiner of English *Second Paper*. He has been an examiner for 5 years under the Board of Intermediate and Secondary Education,

Dhaka. He informed that he enjoyed the task of scoring. He said he also did it for some financial benefit. The researcher observed him while marking the scripts. He took on an average 2.5 minutes to examine a script. Like E3, he also underlined the sentences and marked them. When he was asked why he took so little time to evaluate the answer scripts, he replied that it was unnecessary to use longer time for checking a script as he could understand the quality of the answers just reading first few lines. He said, for composition section, the students answered very poorly; they committed spelling mistakes; and no sentence was grammatically correct.

The researcher asked, then, why he gave credit for the wrong and poor answers. He echoed the voice of other examiners saying that he did not want to do it, but the board authority instructed them to be liberal while checking the scripts. He said that the education board advised them to give good credit for any composition item if the student could generate or organize any idea on a given topic. He commented that the present examination system should be changed and modernized.

Findings of the interviews with the 4 examiners drew a picture of present examination and scoring system. One the one hand, the test contents were not fit to test the target language (i.e. English) ability, on the other hand, the scoring/marking system was undependable. If any external pressure or bias influences examiners' scoring, it causes extreme harms for society in general. Vocabulary teaching and testing gets higher priority, but it is considered the 'crux of the backwash problem'. While vocabulary tests may be a quick measure of language proficiency, once they are established as the only form of assessment, the backwash to instruction resulted in the tests becoming a measure of vocabulary learning rather than language proficiency. Smith (1991b) summarises that testing considerably reduces learning time, narrows the curriculum, and discourages teachers from attempting to meet other goals or from using materials not compatible with formats used by the test makers. She put it, "multiple choice testing leads to multiple choice teaching" (p.10).

Negative washback occurs when the test items are based on an outdated view of language, which bears little relationship to the teaching curriculum. Similarly, Wall and Alderson (1993) reason that if the aims, activities, or marking criteria of the textbook and the exam contain no conflicts and the teachers accept and work towards these goals, then this is a form of positive washback. Negative washback

would be evidenced in the exam having a distorting or restraining influence on what is being taught and how. Swain (1985) commented that "It has frequently been noted that teachers will teach to a test: that is, if they know the content of a test and/or the format of a test, they will teach their students accordingly" (p. 43). The findings of interview suggested that the reliability of scoring was highly inconsistent. So, the results from the HSC examination in English could not give a true picture of language ability. The interview results also revealed that the present HSC examination in English suffered from required validity and reliability, and it influenced teaching and learning negatively.

5.4.3 Interviews with the Curriculum Specialists

The present researcher conducted in-depth interview with curriculum specialists. The purpose of conducting interviews with the curriculum specialists was to explore the intended curriculum, and particularly the rationale behind the proposed language policies. Some confidential and ethical reasons the participants preferred to be anonymous. The curriculum specialists were interviewed through a semi-structured questionnaire (Appendix-4C). They were working at NCTB. The interview was taken in the form of informal discussion at a table held at NCTB. The researcher asked semi-structured question to elicit required information.

5.4.3.1 The Interview Protocols

The instrument used in interviews with curriculum specialists followed what Patton (2002) refers to as the interview guide approach. In this approach, the researcher listed the questions to be explored in an interview and used the list as a guide to "ensure that the same basic lines of inquiry are pursued with each person interviewed" (p. 343). The researcher did not have to follow these questions in any particular order during the interview (Johnson & Christensen, 2004). Rather, this interview guide provided the topic dimensions associated with syllabi, textbooks, and tests within which he was "free to explore, probe, and ask questions that will elucidate and illuminate that particular subject" (Patton, 2002, p. 343). The Figure 5.93 displays the dimensions of interview questions:

Figure 5.93 Dimensions of the interview with the curriculum specialists
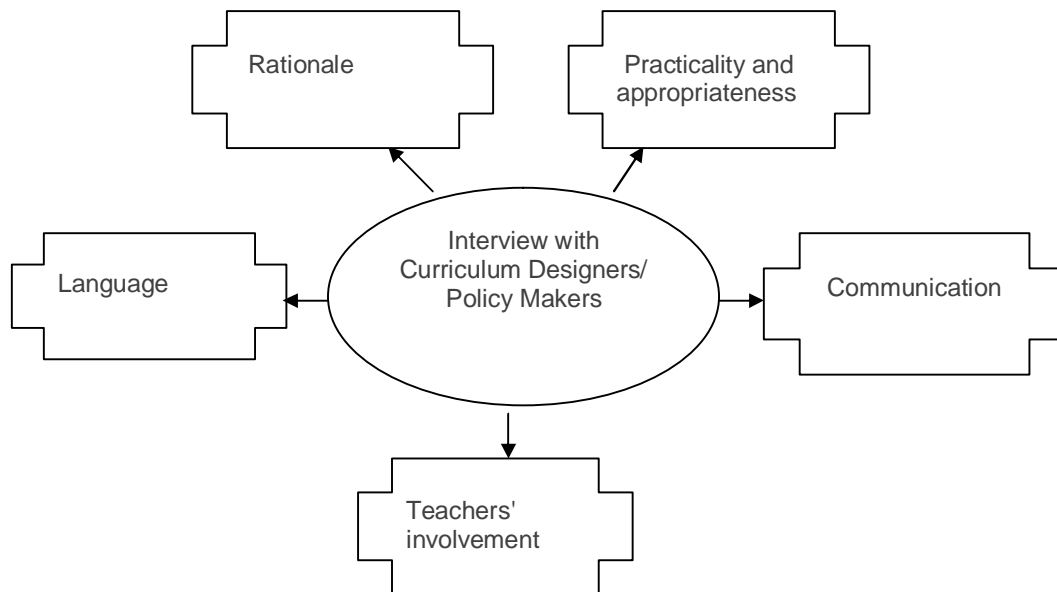
The advantage of this "semi-structured and in depth interview" allowed the researcher to come to the interview with guiding questions and meanwhile remain open to "following the leads of informants and probing into areas that arise during interview interactions" (Hatch, 2002, p. 94).The interview protocols with the policymakers were divided into 5 dimensions as illustrated in Figure 5.93.

The first dimension included the rationale for the design of the syllabus, the compilation of the textbooks and the changes made to the curriculum. The second dimension of the interview questions dealt with the appropriateness and practicality of the syllabus, textbooks, and tests, and the underlying principles of their design. In this case, the present researcher intended to look at the clarity of the syllabus, how the textbooks actually covered the syllabus, and for what purposes the HSC test results were used. The third dimension of the interview questions focused on the communication strategies of each decision-making organization in disseminating the syllabus and textbooks, and on the extent to which EFL teachers' fidelity to implementation was expected. In addition, the researcher wanted to discover participants' perspectives regarding the impact of the HSC examination in English on teachers, students, and the college itself. The fourth dimension of the interview questions was geared to the EFL teachers' involvement in the development of the syllabus and textbooks. The fifth dimension dealt with teachers' language

359

proficiency in delivering the curriculum, and students' English proficiency and the consequences of their failure in the HSC English test.

5.4.3.2 Results of the Interviews

The researcher asked the curriculum specialists to make comments on the present EFL education at the higher secondary level in Bangladesh. Accordingly, they commented that the quality of English education had been deteriorating during the last few years though huge quantitative improvement had been made, the overall pass rate was high, and numbers of GPA-5 had been increased. The curriculum specialists remarked that the quantitative improvement remarkably noticed after the communicative language teaching had been (CLT) introduced in 2000. One of the curriculum experts pointed out that the examination system and test contents were reformed in 2003 in favour of the students. All three interviewed curriculum designers agreed that writing the textbook, *English for Today* for classes 11-12 was a great achievement. They also believed that the book, *English for Today* could not be utilised properly due to lack trained teachers. The interviewed experts suggested that the new EFL syllabus (introduced in 2000) and curriculum seemed to be highly appropriate and well-suited for EFL education in Bangladesh context. Yet they expressed their frustration saying that unfortunately the goals of the syllabus and curriculum were not still achieved.

Regarding the present examination system and test contents, two of the curriculum experts expressed their dissatisfaction with the examination system and the HSC test contents. They remarked that the test was poorly designed and difficulty level was not maintained. All three experts suspected the validity and reliability of HSC examination in English. They focused that in the present evaluation system, one could score very high without achieving proficiency in English. They also categorically commented that the government's target was to increase the pass rate. They believed that the inappropriate language test hampers classroom education. They informed the researcher that new "National Education Policy 2010" would be implemented step by step. They also suggested that the government had a plan to introduce 'creative question' format in all subjects at the higher secondary level.

The interviewed specialists clearly stated that language teaching and learning in the classroom should centre on students, should reduce teachers' speaking time, and should encourage student participation. Therefore, the experts suggested that teachers employed flexible and practical methods according to where different learners were at different stages. In this sense, policymakers are justified in recommending such an approach for EFL teaching methods in Bangladesh context. The experts emphasized that in order to implement the textbook designers' teaching principles; teachers must both adopt a learner-centred approach and use English entirely in instruction. They contended that teachers' positive change in perception was more important than their language proficiency.

This section has presented the key findings of the interview with the three curriculum specialists. The results revealed five themes regarding EFL education at the higher secondary level, and expected implementation in relation to syllabi, textbooks, and tests. To sum up, all the participants agreed that the intended curriculum had not been formulated for the purpose of being concrete and specific, but rather as a more general guideline functioning purely as symbolic guidance to teachers' classroom instruction. They expected that teachers (as the main implementers of curriculum) would achieve the desired goals of the syllabus and curriculum.

5.5 Conclusion

A high-stakes test has been recognised and used as a tool for educational change for many decades. The findings of this research have confirmed the powerful influence that a high-stakes test can have on teaching and learning. This chapter has dealt with the presentation and discussion of both quantitative and qualitative findings of the different sets of the data obtained from a mixed-methods approach. The research supports the claim that washback is neither simple nor direct, but circuitous and complicated (Wall & Alderson, 1993). The quantitative and qualitative results have been discussed in conjunction with the research questions posed in the present study. A close examination of different data sets has revealed some recurring themes with regard to the different factors involved in the washback phenomenon.

This chapter has discussed the findings in light of the theoretical framework presented in Chapter Two. The different types of data, qualitative and quantitative, are intended to serve different purposes. The qualitative data collected through interviews and classroom observations has allowed for a more intensive and in-depth examination of how individual teachers, examiners, and other stakeholders (e.g. curriculum specialists) have perceived and reacted to the HSC examination in English with respect to the specific teaching contexts. As is previously presented in detail in Chapter Four, the qualitative data are supplemented with the survey data. The surveys were used, for it is assumed to be well- suited for quantifying the qualitative data and providing descriptions and comparisons of patterns of both the students' and the teachers' behaviours and beliefs in teaching, learning and testing. The value of this set of data is twofold. On the one hand, they have enhanced the understanding of the participants' interpretations of test impact, their conceptions of teaching, and the meanings they have attached to the classroom practices. On the other hand, they have provided a clear picture of how the teachers' thoughts and conceptions are related to their activities and practices observed. This chapter consists of four sections (see 5.1, 5.2, 5.3 and 5.4); each section has presented and discussed the findings derived from particular types of instruments.

Section 5.1 has presented and discussed the findings obtained from the questionnaire surveys (student survey and teacher survey). The questionnaire responses showed the popularity of test practice activities. The findings of the questionnaire surveys revealed that the HSC examination was driving teaching and learning in the EFL classroom. The phenomenon was interpreted as negative by many researchers (e.g. Cheng, 2004) saying said that assessment-driven teaching was leading to loss of teaching and learning time. Narrowing the curriculum by leaving out what was not tested was also evident and seen as a negative washback. The materials and teaching methods were also negatively influenced by the examination. Test related commercially produced materials were used by most of the teachers and students inside and outside the classroom. The study has identified some other factors that are encouraging the teachers to teach to tests, such as time constraint, and pressure from students, their parents and college management.

Section 5.2 has reported the findings and discussion resulted from the classroom observations. From a methodological perspective, the intensive classroom

362

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

observations provided this study with rich data. However, the inclusion of several additional sources of data may have contributed to the observations. This section has focused on the findings of observing the teachers and students in the classroom setting about their lessons and opinions on the HSC examination. The COLT, UCOS and self-made checklist recorded various classroom activities to examine whether the activities of the EFL teachers were driven by the HSC examination. It was found that most of the observed classes (80%) were highly teacher-controlled. It seemed that the teachers still dominated most of the classroom interactions, and that teacher talk was the predominant strategy in their classes. The activities mainly emphasised reading skills, vocabulary building, language points and sentence structures, which limited students to producing isolated sentences, and limited them to be assessed merely for grammatical accuracy. Moreover, the nature of teacher talk, as a percentage of class time, was similar in most of the teachers' cases. The detailed findings suggested that the negative washback effect had driven the attitudes, behaviours, and perceptions of the teachers. The findings revealed that more than 80% teachers were entirely reliant on test related commercially produced materials. Washback was seen to occur quickly and strongly in the creation of teaching those materials. The perceptions of the students and teachers regarding teaching and learning activities were also directly influenced by the HSC examination. The findings suggested that the teachers pronounced the name of the HSC examination several times during the lesson, and reminded the students that the examination was not far away. The detailed findings of the classroom observations suggested that there was washback effect of the HSC examination on teaching and the classroom behaviours.

Section 5.3 has documented the findings obtained from the examination related documents analyses (syllabus and curriculum, textbook analysis, HSC test/question papers analysis, and HSC answer scripts analysis). It was found that the syllabus and curriculum set specific goals and objectives, and covered the needs of the language learners. The findings of the textbook analysis discovered that *English for Today* (for Classes 11-12) directly corresponded to the HSC English syllabus and curriculum, and was well-suited one to enhance EFL teaching and learning at the HSC level. The findings focused that the HSC EFL test extremely suffered from lack of validity and reliability as a language test because the examination contents

and items could not assess the students' communicative competence. The findings also depicted that the scoring/marking procedures of the answer scripts were unreliable. There were no scoring guidelines for the examiners. The scores obtained by the testees (students) varied from examiner to examiner to a large extent (up to 20%). It exerted negative impact on classroom teaching.

Section 5.4 has offered the presentation and discussion of the findings received from the interviews with EFL teachers, EFL examiners and curriculum specialists. The findings disclosed that the EFL teachers used most of the class time for the test preparation. The teachers usually used commercially produced test related materials, and preferred to use the mother tongue Bengali as a medium of instruction. They were negatively influenced by the HSC examination which was the evidence of negative washback of the test. The examiners of English believed the HSC examination, results were not the true indicators of the students' proficiency level in English. The findings of the interviews with the curriculum specialists suggested that the English syllabus and curriculum designed 2000 should have been revised. They also emphasised the modification of the HSC EFL test contents and items to assess the communicative competence of the students.

Thus, in every section, the researcher has tried to examine how the findings revealed the evidence of washback of the HSC examination. He has also sought the answers to the research questions through discussions. He has examined and re-examined the findings by cross-referencing with the findings obtained from other instruments. References of previous studies and expert opinions have also been presented in support of the findings of the present study. The next chapter, Chapter Six, summarises the findings, offers answers to the research questions, touches upon the implications of the findings, and puts forward recommendations. It also points out some directions for further research.

Chapter Six
Conclusion

This chapter first summarises the findings of the study. Then it puts forward the answers to the research questions posed so as to carry out the study. Later, it deals with the implications of this investigation. Based on the findings, the researcher has made some recommendations for enhancing EFL education including the examination at the HSC level. The researcher also proposes a washback model which demonstrates the role of different stakeholders, factors and beliefs in mediating positive washback. The chapter continues with suggestions for further research. Finally, it draws conclusion of the thesis based on the findings of the study.

6.1. Findings of the Study in Brief

The study applied the MMR approach to data collection and data analysis. The analysis and discussion of findings of the study are presented in the previous chapter (Chapter Five). In this chapter, the major findings of the study are summarised and presented by themes.

The results, obtained from both qualitative and quantitative data, suggest that the HSC EFL examinations have a strong negative washback on the classroom instructions and other related areas of EFL education such as the syllabus and curriculum, teaching content, teaching methodology, EFL learning, etc. The study has discovered a number of interesting findings that are cross referenced through a number of ways. The results have also revealed that the depth, extent and direction of the effect differed with the affected areas. The contents of teaching seemed to be the area showing changes directly triggered by the test. This was in line with the results of previous studies on washback (e.g. Wall and Alderson, 1993; Alderson and Hamp-Lyons, 1996; Cheng, 1997) which found the content of language teaching as the area readily susceptible to changes as a result of tests.

The use of COLT and UCOS in combination with the further specific analysis enabled the examination of the communicative orientation of classrooms.

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

365

This was an attempt not only to determine the range of activities that might occur in the HSC exam preparation class, but also to identify the amount of lesson time in which the students in the 10 observed classes were actively communicating, as this would be an indication of good classroom practice which could in turn possibly be seen as a result of a 'good' test. Teachers' and students' perspectives were elicited and cross-referenced to the findings of the instruments, using a combination of purpose built questionnaires and interviews. The combination of the instruments used in the study allowed for a meaningful comparison of findings.

The key intent of this study was to investigate washback within the context of HSC public examination in English in Bangladesh. The present research was an intensive study which applied a number of data collection instruments. The study began by surveying the literature to try and find a clear definition of washback and its influence on various aspects of language teaching and learning. Considering previous discussions and research, a schedule for the research was designed to include the perspectives of the teachers and their students, as well as detailed classroom observations. Having observed 10 EFL teachers, the present researcher have analysed the collected data and allowed any pattern to emerge. After the classroom observation had been over, the present researcher conducted post-observation interview with the observed teachers.

As outlined in the previous chapter (Chapter 5), this study provides information about the nature and variability of washback. Owing to the various factors and contextual factors underlying the washback phenomenon, the EFL test at the HSC level seems incapable of helping teachers initiate a deeper self awareness of teaching to generate a positive washback on language instructions and students' learning. It is not capable of effecting fundamental teacher conceptual and behavioural change. The results seem to suggest that seeking immediate change in teacher beliefs, knowledge, and behaviour is impractical, which confirms other findings (e.g. Freeman & Johnson, 1998; Richards, 2008; Woods, 1996) that claim that there is often little immediate evidence for change in teachers' practice(s). Drawing on the theoretical framework outlined above (in Chapter Two), the process of initiating teacher change is complex and developmental, and it may develop and emerge over time through the process of teaching and training.

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

366

There is also adequate evidence showing that the HSC examination in EFL has had a different impact on different teachers, since great variation was noted in their interpretations of and reactions to washback. The results reveal that while the participants appear to have interacted with institutional contexts (or context of work) in different ways, they seem to respond and react to washback in accordance with the rules and norms that prevail in their institutional contexts. Thus, a conclusion that can be drawn is that both personal teacher factors and institutional contextual factors are involved in the process of washback. In addition to providing information about washback, this research has provided information about how a number of factors are manifested in the washback process.

One important finding is that the interconnected attributes seem to form a core that threads itself through the process of teaching as well as the process of washback. The limited washback in this study might be attributed to the combined reasons of teachers' misconceptions of teaching and learning, inadequate knowledge, insufficient in-service training, and guidance for change. Teachers' pedagogical knowledge was found to affect teacher practice in a deterministic way.

Consistent with the findings of Andrews (2004), Shulman (2000), and Turner (2008, 2009) that view teacher pedagogical knowledge as an essential component of teacher professionalism, this study has shown that this dimension of knowledge exerts a strong influence on how teachers interpret testing innovation and change their practice. This finding also reinforces Richards' (2008) assertion that such knowledge contributes to the formulation of teachers' working principles that guide their teaching behaviour and functions as the source of teachers' practices.

It is discussed that Bangladesh has an educational context which is centralised, knowledge-focused, and examination-based. It also has a socio-cultural environment in which teachers have to cope with large classes and a lack of freedom to choose their own textbooks and content of teaching. Classroom observations have shown that the traditional outdated teaching norms typical of the Bangladesh educational context are still present. The participating teachers in this study were found to hold a knowledge transmission perspective and to adopt transmission-oriented teaching styles.

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

367

They also viewed students as passive recipients of transmitted knowledge rather than active participants in the construction of meaning. It appears that the Bangladesh educational and socio-cultural environments may have a part to play in nurturing the above-mentioned teaching patterns and norms. Thus, it was imperative to adopt a socially and institutionally situated view about teacher learning and knowledge and consider the effects of both social and institutional contexts upon teacher practice. As pointed out by other language testing researchers (e.g. Turner, 2005), washback cannot be understood apart from the socio-cultural environments in which it is embedded. Based on the findings of the present study, the researcher supports the very authentic view of Turner (2005). Now, the findings of the study are summarised below by themes.

6.1.1. Findings Related to the Syllabus and Curriculum

Washback has deep relation with the syllabus and curriculum. Test contents can have a very direct washback effect upon teaching curricula. This study reviewed more than one hundred qualitative studies to interrogate how high-stakes testing affects the curriculum, defined here as embodying content, knowledge form, and pedagogy. The findings of this study complicate the understanding of the relationship between HSC (high-stakes) EFL testing and classroom practice by identifying contradictory trends. The primary effect of this high-stakes testing is that curriculum content is narrowed to tested subjects, subject area knowledge is fragmented into test-related pieces, and teachers increase the use of teacher-centered pedagogies. However, this study also finds that, in a significant minority of cases, certain types of high-stakes tests have led to curriculum content expansion, the integration of knowledge, and more student-centered, cooperative pedagogies. Thus, the findings of the study suggest that the nature of a high-stakes test such as the HSC examination, induced curricular control is highly dependent on the structures of the test itself. The other major findings on the syllabus and curriculum are presented below:

1. Maximum teachers and students are not aware of the objectives of the HSC syllabus and curriculum.

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

368

2. Though most of the teachers and students are of poor curriculum knowledge, they believe that present the syllabus and curriculum can help the students learn English language.

3. It is found that the teachers do not teach every section and lesson of the prescribed syllabus. They teach those lessons which may be tested in the examination.

4. Maximum teachers and students confirm that they do not care about the syllabus while preparing for the examination. Most of them follow the 'hidden syllabus' (Caine, 2005) such as past question papers, model test papers, etc.

5. It is found that the textbook correspond to the EFL syllabus; the whole syllabus is incorporated in the present textbook English *for Today* for classes 11- 12. The findings show that the respondents feel pressure to cover the syllabus in stipulated two years time.

6. It is found that all items of the syllabus are not tested; some of the items remained untouched and ignored since the publication of the book in 2001.

7. It is found that washback of the HSC examination in English influences the academic behaviour of the teachers and learners.

6.1.2. Findings Related to the Textbook Materials

Washback affects various aspects of teaching and learning which can be categorized as follows: stakeholders, testing and mediating factors, learning outcomes, syllabus and curriculum, materials, teaching methods, feelings, attitudes, and learning, etc. The study finds washback of the HSC EFL test affects the textbook materials; the teachers mostly use those materials that are directly related to their test. Existing washback models/hypotheses contend that tests alone or at least for most of the part trigger the perceived washback effects, but empirical studies show that both testing and mediating factors play essential roles involved in the mechanism of washback effects.

Another significant finding is that both learners and teachers bring model test papers, guidebooks and past question papers in stead of the EFT; they solve the

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

369

model questions one by one in the class. Test preparation classes might increase students' scores, but the score gains are not always statistically significant. Moreover, class instruction of exam-specific strategies and non-class instruction factors such as students' initial proficiency, personality, motivation, confidence, and exposure of environment all possibly contribute to a score gain.

Another interesting finding is that there are so far 20 test–related commercially produced are found available in the market. The publishers and distributors visit the colleges and distribute the test-related book among the EFL teachers at free of cost for publicity of their book. Andrews, et al. (2002) also speak of the large role played by published materials in the Hong Kong classroom, citing a previous study by Andrews (1995) in which the teacher respondents were found to spend an estimated two-thirds of class time working on exam-related published materials. It is seen that exam materials are heavily used in classrooms particularly as the exam approaches. Some more findings are given below:

1. Though the authority promised to formulate a "Teacher's Guide" but it is not yet written; therefore, the teachers, especially the novice, face much problem in teaching the text communicatively.

2. A considerable number of teachers do not understand the philosophy/ approach of the textbook.

3. All the teachers seem willing to go along with the demands of the exam (if only they knew what they were).

4. While teaching, the teachers do not communicate the lesson objectives with the learners.

5. It is found that the textbook contains sufficient exercises and opportunities for practising EFL; it is considered as a well-suited textbook; the topics are interesting, yet the learners do not study it seriously.

6. The textbook analysis finds that the textbook contains sufficient exercises for communicative language practice such as pair work, group work, conversation practice, drills, dialogues, etc.

7. The teachers skip certain lessons and exercises of the textbook because they are less likely to be tested.

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

370

8. Both the teachers and students heavily rely on the test related materials such as test papers, past questions, and model questions in the EFL classes.

9. The students and teachers believe that if the whole textbook is taught and studied, the students will perform badly in the examination.

10. Thus, the study finds that the washback influences the use of materials in the class towards teaching to the test.

6.1.3 Findings Related to the Teaching Methods and Approaches

The study finds that the teaching method varies teacher to teacher. The EFL syllabus and curriculum desired that the teachers would apply CLT in the class for communicative language teaching; but still, they use the grammar translation method. The mother tongue is vastly used as a medium of instruction. The teachers translate the English text into Bengali, and transmit the knowledge thematically, not from linguistic point of view. Watanabe (2004b, p. 44) notes that 'students rarely asked questions even during exam preparation lessons'. Cheng (1998) points out that while teachers talk less to the whole class as a result of the revised exam, the teacher talking to the whole class remains the dominant mode of interaction.

1. Many teachers are unable, or feel unable, to implement the recommended methodology. They either lack the skills or feel factors in their teaching situation prevent them from teaching the way they understood they should.

2. Many of the teachers do not like that the students ask any question. It is found that most of the teachers do not encourage and motivate their students to speak English and ask questions.

3. The blending of English and Bengali hinder the practice of target language.

4. Inadequate or no training and teachers' professional backgrounds lead to unchanged methodologies because they don't know how to change, not that they do not want to change.

Another notable theme that has emerged from the findings is that the 'learner factor' is considered by teachers to play an essential role in their decisions concerning teaching methodology. In general, teachers blame tests, learner factor, large class, etc. As recounted by the participants of this study, in addition to the

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

371

impact of the HSC examination, the 'learner factor' also constitutes a constraint on their teaching. Based on their accounts, the factor chiefly involves: 1) learner beliefs, 2) learner proficiency, 3) learner habits, etc.

6.1.4. Findings Related to the Classroom Tasks and Activities

All the classes observed were found to consist predominantly of materials written for language students; contained a significant number of practice tests; included exam-related activities and incorporated few academic study skills. Four of the classes were found to be examples of a more traditional approach to test preparation, which focused on familiarising students with the test and providing opportunities for test practice both in and out of class.

It is found that majority of the teachers tended to "teach to the test." It is found more test-related activities (e.g. offering test-taking tips, doing question analysis) in the exam preparation classes. The observed teachers pointed out that they had to change their curricular planning and instruction to a certain extent in order to meet the testing objectives.

The analysis with the UCOS instrument showed that all the teachers spent similar amounts of class time getting their students to analyse their performance on tasks and to identify the areas they needed to improve in:

1. test-taking is much more common in EFL classes;
2. teachers teach test-taking strategies;
3. teachers skip language classes to study for the test;
4. teachers talk more and students have less time available to talk in EFL classes;
5. teachers occupy most of the class time;
6. there is less turn-taking and turns are somewhat longer in HSC EFL classes;
7. less time is spent on pair work in classes;
8. the HSC examination is referred to many times in EFL classes;
9. Bengali is used much more in EFL classes;
10. there is no laughter in EFL classes;
11. the teachers give model test as a means of test preparation;

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

372

12. students practice items similar in format to those on the test;

13. students study vocabulary items and grammar rules; and

14. students read widely in the target language.

6.1.5 Findings Related to the Practices of Language skills

Washback of the HSC examination in English affects the practise of integrated skills of English as a foreign language. The questionnaire survey as well as the direct classroom observation finds that two major skills of EFL: listening and speaking are hardly practised in the class. The post- observation interview with the observed teachers suggested that this is waste of time. They said speaking and listening practices will bring no good for the students. Three teachers informed that their students did not prefer to practise, even though they tried to teach them listening. A number of teachers, however, consistently skip over the listening lessons in their textbooks, because they know that listening will not be tested in the exam. Other teachers may 'do listening', but in a way that does not resemble the textbook designers' intentions.

6.1.6 Findings Related to the Teachers' and Students' Academic Behaviours and Beliefs

It is found that the effect of HSC examination in English changes the academic behaviour of both teachers and students. The teachers and learners are now within the circle of test boundary. In addition, teachers' beliefs and attitudes regarding the immediate goals of teaching and their own limited ability to use the language effectively contribute to their being unable to affect the positive changes (a shift in English language teaching to a more communicative orientation) the test developers intended to create.

The findings summarised above have implications for both language testing research and other types of research (e.g., general education, language education, and educational innovation), for the findings of this study share a number of similarities to those of other research. For instance, one of them is that they all aim at improving teacher practice, classroom teaching, and communicative competence. Another one is that they share a common focus on the 'teacher factor' or 'teacher

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

373

role' in the classroom. To be specific, the teachers are seen as the major change agents who play a central role in shaping classroom events (Borg, 2006; Davison, 2008; Davison & Hamp-Lyons, 2008; Turner, 2008, 2009).

The following section discusses the implications of these findings. The areas most influenced by washback are found to be those related to immediate classroom contexts: teachers' choice of materials, teaching activities, learners' strategies, and learning outcomes. The study also reveals that non-test-related forces and factors operative in a given educational system might prevent or delay beneficial washback from happening. Based on the theoretical assumption underlying the definition of washback adopted in this study, many consequences which cannot be traced back to the construct of the test are outside the limits of a washback study.

6. 2. Answers to the Research Questions

In order to facilitate the investigation of washback of the public examination on teaching and learning EFL at the HSC level, the researcher formulated the following research questions. In this section, the research questions of the present study are answered. A research question is the methodological point of departure of scholarly research in both the general education and applied linguistics. The research answers any question posed. The research question is one of the first methodological steps a researcher has to take when undertaking research. The research question must be accurately and clearly defined. Choosing a research question is the central element of both quantitative and qualitative research and in some cases it may precede construction of the conceptual framework of study. In all cases, it makes the theoretical assumptions in the framework more explicit, most of all it indicates what the researcher wants to know most and first.

Based on the research purposes, the study looked at the washback effect of the HSC public examination on the EFL teaching and learning both at the macro level with respect to major parties within the Bangladesh educational context and at the micro level with regard to different facets of classroom teaching and learning. Therefore, the study set up the following research questions to answer:

RQ₁. Does washback of the HSC examination influence EFL teaching and learning?

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

374

RQ₂. Does the HSC examination have any washback effect on the syllabus and curriculum?

RQ₃. To what extent does the test content influence teaching methodology?

RQ₄. What are the nature and scope of testing the EFL skills of the students at the higher secondary level?

RQ₅. What are the effects that an examination preparation process can have on what teachers and learners actually do?

RQ₆. What is the effect of the HSC examination on the academic behaviour, feelings, perception and attitudes of teachers and students?

6.2.1 Answer to Research Question 1 (R₁)

RQ₁. Does washback of the HSC examination influence EFL teaching and learning?

This is a leading research question posed to the study. As mentioned, an MMR approach was applied to data collection and data analysis. To investigate the washback effect, data was collected from a number of sources and a number of ways such as test analysis answer scripts evaluation, in-depth interview with the examiners, direct classroom observation, post observation interview, and questionnaire survey. One important dynamic of washback revealed in the present study reflected test anticipation in general.

Throughout the study, different forms of washback behaviours were observed on teaching and learning. From the classroom observation, it was noticed that teachers frequently used the name of the HSC examination; they were advised their students to prepare themselves; they were found using commercially produced test related materials; they hardly used target language in the class (sections 5.2.1 to 5.2.7).

From the Questionnaire survey, both the teachers and students replied that EFL examination system hindered their teaching and learning. The teachers taught those items that are directly related to the examination, and skipped certain sections that are not tested. The finding of the Q11, Q12, Q13, Q15, Q27 Q28, Q29, Q30 and Q31 answered the present research question. The findings were discussed in the Chapter 5.1 (sections 5.1.3.2.3 to 5.1.3.2.5 and 5.1.5 to 5.1.5.1.2). All the findings

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

375

proved that the HSC examination in English influenced classroom teaching and learning negatively.

The findings from the in-depth interviews with the EFL teachers revealed that the teachers believed they should not have taught those items that were not tested. They also considered it simply waste of time, and did not want to take any risks teaching that item. Throughout the whole study, a number of observable evidences were found to answer that washback existed on classroom instructions in general, and the washback negatively impacted. Positive washback can be used to influence the language syllabus and curriculum. As Davies (1990) mentions, washback is inevitable and it is foolish to pretend that washback does not happen. Therefore, in order to prepare students for the examination, the communicative way of teaching will be adopted in our classes and this positive washback helps us change the curriculum the way we want.

Positive washback would result when the testing procedure reflects the skills and abilities that are taught in the course, as, for instance, with the use of an oral interview for a final examination in a course in conversational language use. Therefore, when there is a match between the activities used in learning the language and the activities involved in preparing for the test, it can be said that the test has positive washback. As Brown (2002) states washback becomes negative washback when there is a mismatch between the content (e.g., the material/ abilities being taught) and the test. Tests narrow down the curriculum, and put attention to those skills that are most relevant to testing. The next part shows washback works on the syllabus and curriculum.

6.2.2 Answer to Research Question 2 (R$_2$)

RQ$_2$. Does the HSC examination have any washback effect on the syllabus and curriculum?

The answer to this research question was found in Chapter 5.1.2 (sections 5.1.2.1.1 to 5.1.2.3.2). A number of statistical analyses of data were conducted to draw the results. The findings of a series of questions (Q1-Q7) reflected that teachers did not teach all the contents of the syllabus and curriculum due to test pressure; teachers gave little attention to the communicative competence. Positive

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

376

washback takes place when tests induce teachers to cover their subjects more thoroughly, making them complete their syllabi within the prescribed time limits.

A creative and innovative test can quite advantageously result in a syllabus alteration or a new syllabus (Davies, 1990). An examination achieves the goals of teaching and learning, such as the introduction of new textbooks and new curricula (Cheng; 2005). Both groups of respondents confirmed that they were not even aware of the objectives of the syllabus and curriculum; on of the reasons of their ignorance is that the teachers taught to the test, not the syllabus contents. They narrowed down the syllabus and curriculum for the immediate benefit of test preparation. Therefore, it may be concluded that the HSC examination in English has negative washback effect on the syllabus and curriculum.

6.2.3 Answer to Research Question 3 (R₃)

RQ₃. To what extent does test content influence teaching methodology?

A test will influence how teachers teach; and a test will influence how learners learn (Wall & Alderson, 1993). But their research in Sri Lanka found no evidence of methodology change. While Alderson and Wall (1993, p. 127) said that their Sri Lanka study showed the exam 'had virtually no impact on the way that teachers teach'. Watanabe (1996) find large differences in the way teachers teach towards the same exam or exam skill, with some adopting much more overt 'teaching to the test' , 'textbook slave' approaches, while others adopted more creative and independent approaches (p, 292).

Some researchers (e.g. Green, 2006, 2007; Johnson, 1992; Nyawaranda, 1998; Sato and Kleinsasser, 1999; Tan, 2008; Turner, 2006, 2008; Wang, 2008; Cheng, 2004) found that washback influenced the teaching methodology in varied degree. Teacher beliefs are consistent with their prior experience and instructional approaches. There is, therefore, an increasing realisation in the field of assessment that the "teacher factor" is fundamental to the kind of washback effect that takes place in the classroom.

The present researcher found that washback influenced the teaching methodology. A group of questions (Q18-Q26) were asked to the respondents on the methodology issue. The findings of those questions are presented in the Chapter 5.

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

377

1.4 (sections 5.1.4.2 to 5.1.4.2.5). Data were collected on a number of issues to observe any change in the EFL classes. The findings were cross-referenced and examined with other questions for the reliability of the findings. The different classroom actions of the respondents proved that washback of the HSC examination influence teaching and learning. These results are also validated by the classroom observation results. During the observation, the researcher found that most of the teachers (except T1 and T6) did not apply the CLT in the classroom; they used the mother tongue as the medium of classroom instruction and translated the text into Bengali, their mother tongue.

The evidence of washback influence on teaching methodology was further proved by the results of in-depth interview with the observed teachers. In replying to the same question, most of the teachers (6 out of 10) said they taught in a way that their students could understand better and they did not consider whatever method it was. The teachers ignored the spirit of CLT, and aims and objectives of the EFL curriculum.

6.2.4 Answer to Research Question 4 (R$_4$)

RQ$_4$. What are the nature and scope of testing the EFL skills at the higher secondary level?

The present study found that there was no scope at all for testing listening and speaking skills in the present examination system. The grammar, vocabulary, some language elements, reading comprehension, and writing skills were tested in the examination. The survey of Q6, Q26, and Q33-Q37 asked about the practice of the foreign language skills. The findings suggested that listening and speaking were neither tested nor taught.

The analysis of the HSC EFL test found that there were no scope of testing communicative competence, particularly testing of listening and speaking were ignored. These findings were further validated by the results of answer scripts analyses (section 5.3.4). The reality of not testing all the skills of a foreign language further admitted while interviewing with the examiners and observed EFL teachers (sections 5.4.1.1 to 5.4.2.4). It is now well grounded fact that two major skills are not at all tested, and consequently teachers feel uninterested to teach those skills.

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

378

This is a very strong observable evidence of negative washback on the classroom practice of all the foreign language skills adequately.

6.2.5 Answer to Research Question 5 (R₅)

RQ₅. What are the effects that an examination preparation process can have on what teachers and learners actually do?

The learners are the key participants whose lives are most directly influenced by language testing washback. The washback influences the test takers directly by affecting language learning (or non-learning), while the influences on other stakeholders will affect efforts to promote language learning. The test-takers themselves can be affected by: the experience of taking and, in some cases, of preparing for the test; the feedback they receive about their performance on the test; and; the decisions that may be made about them on the basis of the test.

Shohamy et al. (1996) contend that an important test promotes learning, while Cheng (1998) shares a similar finding by saying that tests motivated students to learn but that their learning strategies did not change significantly from one test to another. A recent study by Stoneman (2006), investigated how students prepared for an exit exam in Hong Kong. The result showed that students were motivated more and spent more time in preparing for higher-status exam (IELTS) than the lower-status test (GSLAP), but preparation methods were much the same. The examination preparation activities can influence negatively and positively. A group of questions (Q27- Q32) was to obtain the information about examination preparation activates of the teachers and students. The previous chapter (sections 5.1.5.1 to 5.1.5.3) presented the finding of those questions.

Many research studies reveal that a test affects participants, processes, and products in teaching and learning. Students, teachers, administrators, material developers and textbook writers may be included under the term 'participants'. Tests have impact on the lives of test takers, classrooms, school systems and even whole societies (Hamp-Lyons, 1998). From the above discussion, it can be concluded that the examination preparation process influences both the learners and teachers, and their actions sufficiently.

6.2.6 Answer to Research Question 6 (R$_6$)

RQ$_6$. What is the effect of the HSC examination on the academic behaviour, feelings, perception and attitudes of teachers and students?

Their perceptions, academic behaviour, feelings, and attitudes towards their work are likely to be affected by a test. As mentioned, washback affects various aspects of teaching and learning, such as the syllabus and curriculum, stakeholders, materials, teaching methods, testing and mediating factors, learning outcomes, feelings, attitudes, and learning, etc. The effect of the high-stakes test generates huge pressure for both learners and teachers.

The students feel that the results of the examination will shape their career. They feel embarrassed if they fail in the examination. Most of the learners suffer from anxiety and tension for the examination. Most often they feel pressure from parents to make good results. Thus, their academic behaviours are test-driven; they believe there is possibility of better career. If some students fail in the examination they are insulted and humiliated for all quarters. Most of the teachers also feel pressure from their authority to make the exam results better; the teachers believe that they get feedback from the exam results of their students. Both groups of respondents believe that teaching and learning would take place better if there was no pressure for good results. All the findings prove that negative washback effect influences their (student and teachers) academic and, to a large extent, personal life.

6.3 Implications of the Study

The study shows that washback is a complex phenomenon rather than being a direct and automatic effect. Furthermore, washback exists on a variety of teaching and learning areas (e.g. curriculum, methods of teaching, classroom assessment, student learning, feelings and attitudes of teachers and students); it is also found that washback varied in form and intensity (Cheng 2005) as well as in specificity, length, intentionality and value (Watanabe 2004b). Washback was also found to be broad and multi-faceted, and brought about through the agency of many independent and intervening variables beside the exam itself. These include teachers and students, the status of the subject being tested, resources and classroom conditions, management

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

380

of practices in the schools, communication between test providers and test users and the socio-political context in which the test is put to use.

However, while theory and research have so far succeeded in formulating a general framework for the washback phenomenon that involves several components and intervening factors, there is still a need to account for the interplay between them. Actually, an important step forward would be to construct a conceptual model of washback which would provide a synthesis of its multiple components and demonstrate the interaction between them and the factors involved. This would provide a better understanding of the reasons why washback occurs and an explanation of how tests can be designed in order to engineer positive washback.

The present research began with a very broad question about the effect of the HSC exam. As the research progressed, it became clear that the nature of the exam washback was not as straightforward as the language teachers believed. In fact, the findings suggest that test washback is a very complex process. The study showed that the HSC exam could negatively influence teaching and learning, they could not have exclusive power over what is happening in the classroom. Rather other factors, direct or indirect relation to the exam, seemed to be playing a greater role or a role at least as significant as the exam in shaping the classroom teaching and learning and contributing (or not) towards positive washback. The various factors that need to be taken into account when an attempt is made to induce positive washback of high-stakes tests include the following stakeholders: textbook writers and publishers, teachers, students, schools, parents, local education systems and local society. It is also possible that other factors, like mass media (ELT newspapers, newsletters, TV and radio) might have interfered with the authority's efforts to bring about the intended washback effect, but these were not explored in the present study.

However, the above factors, likely to enhance or interfere with the test's intended purpose, do not function in isolation. The picture that emerged from this study is that the influence of the exam at the classroom level is a dynamic and interactive process between the exam, the textbook, the teacher and students, each influencing the other, in varying degrees of quantity and quality. This interaction is in turn shaped by the interrelation of several other related factors which operate within the immediate and broader context, such as test constructors, material writers and publishers, schools, parents and the local education system. These factors can

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

381

then be seen as additional agents in helping to engineer or hamper the intended washback of an exam.

The present study has examined washback of the HSC examination, a high-stakes exam, taking as its specific focus teachers, exam preparation materials and learners. The results showed that high-stakes tests, powerful as they are, might not be efficient agents for profound changes in an educational context. While the study reiterates the complexity of investigating washback noted by previous studies, it also provides an indication as to the sources of this complexity that can be traced both inside and outside the classroom context.

The present study proposed a model for promoting positive washback (Figure 6.1). In the model, washback is represented as an open loop process identifying the number of stakeholders and factors involved in the process and attempting to portray the relationship between them. However, despite being a multi-directional relationship among stakeholders, the model, in its visual representation below, is 'simplified' to make it possible to represent it graphically. In the model, the nature of exam washback is circuitous and interactive. Exam washback is indirectly engineered on teaching and learning that takes place in the exam preparation classroom through the understanding of the exam requirements by various intermediary participants.

From the findings of the current study, it is found that exam washback is mediated through commercial exam-preparation materials, which are shaped by the perceptions of the needs of teachers and students by writers and publishers. The exam-preparation materials mediate between the exam intentions and the exam preparation class. The teacher's role is also crucial in the process as they mediate between material and students. Within this process, washback is also mediated by the HSC colleges, and strengthened by the perceptions and understanding of various other stakeholders operating in the wider local community, such as parents, as well as by the local educational system and beliefs in the exam and the language tested.

The study also revealed that the role of parents was crucial in the washback process. Washback to parents was mediated both through the colleges (progress reports and parents' meetings) and manifested in behaviour such as exhorting students to work hard. The findings also revealed that, very often, parents affected students' language motivation (e.g. prompting students to take up learning English

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

382

at a young age) and instigated an instrumental disposition towards English. Parents' beliefs about the value of the exam and the language were actually shaped by the local context which recognised the HSC examination as an official language qualification and placed a great deal of importance on learning/mastering the language.

The results of the present study indicate that overcoming the barriers to change is no simple task. It requires the joint efforts of authorities, test developers, students and particularly teachers. Its implications are twofold. On the one hand, the Bangladesh examination authorities and test developers could focus on improving test design to facilitate language teaching and learning. The simple reason, as pinpointed by Muñoz and Álvarez (2010), is that "students' successful performance on assessment tasks greatly depends on how well teachers and test developers design those tasks" (p. 37). As illustrated and exemplified in studies by Colby-Kelly and Turner (2007), Davison, 2008, Davison and Hamp-Lyons (2008), Rea-Dickens (2004), and Saif (2006), considerations and efforts directed at alignment of assessment, curriculum, teaching and learning practices do make a difference.

The results of this study, however, indicated that the HSC EFL test is problematic for a number of reasons. To exercise regulation, it is essential to ensure the validity and reliability of the tests employed. Validity refers to the extent to which a test measures what it is intended to measure and not what it is not designed to measure. If what is to be measured is already controversial, the validity of the test is unlikely to be agreed on. Reliability is essentially concerned with how consistently the test does what it is supposed to do. A common test error is associated with the cut-off score. The indication of mastery of certain language abilities rests on reaching the cut-off score. Where the cut-off score of a standard control instrument should be set, or how to determine the cut-off score accurately, is often subjective (Davies, 1990).

This research suggests that for the successful implementation of assessment innovations, the policy makers need to take into consideration the power of beliefs in their planning and strategies. Understanding what types of beliefs lead to desired outcomes and what types of beliefs are stumbling blocks would be of use to help promote positive washback and minimise undesirable negative washback. For example, for a high-stakes test to help teaching and learning, it seems vital for both

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

383

teachers and students to believe that the test is functioning as assessment *for* learning as well as *of* learning. Possible negative washback effects that the test is failing to motivate students and that teachers and students are overloaded with assessments would be likely to be minimised if the belief that the assessment is interfering with teaching and learning is reduced.

In this study, the HSC EFL Examination tasks were found not to correspond highly to tasks in the language use (TLU) domain. Thus, it would appear beneficial for the English exam constructors to address issues and problems inherent in the test. As the present curriculum places an emphasis on facilitating students' communicative competence, test constructors need to, as Shohamy (1993) suggests, do their job to ensure the construct of test validity and increase the match between the curriculum and the test. To this end, organized efforts need to be directed to more task-based test designs guided by contemporary language testing theories (Bachman & Palmer, 1996; Bailey, 1996). In this way, test papers (question papers) would be designed in such a way that not only knowledge of English, but also the ability to use English, is tested. As Bachman and Palmer (1996) suggested, task characteristics need to be considered in order to ensure that test tasks correspond in specific ways to language use tasks. Evidence needs to be provided to demonstrate that the test score reflects the area(s) of language ability they want to measure.

The results of the study have implications for language testing. Unfortunately, the current language testing system in English classrooms in Bangladesh does not reflect this complex and multidimensional nature of language performance, only measuring the fragmented "knowledge" in English, not the "performance," or "proficiency" of the students. In addition, this study has some implications for English teacher education in the EFL context in terms how to empower them. As noted earlier, being a "native speaker" or having a "native-like" competence is not a realistic goal for most EFL students to acquire English. The researcher also pointed out that when it is only native speakers of English who are projected as qualified teachers and speakers of English in students' learning/testing environments, their linguistic/cultural boundaries and possibilities might be seriously limited.

From the findings of the current research, important implications can be drawn for policy makers who are involved in educational reform. The commonly

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

384

held assumption that intended washback could be brought about just by introducing a new or modified assessment (e.g., Chapman & Snyder, 2000) needs to be altered. Supporting the argument of many researchers in the fields of education and applied linguistics (e.g., Alderson, 2004; Chapman & Snyder, 2000; Fullan, 2001; Shohamy et al., 1996; Wall, 1996; Watanabe, 2004a), this research demonstrated that washback is a complex phenomenon in that it appears to be mediated by certain types of contextual factors and various beliefs held by teachers and students. It is clear that intended washback cannot be promoted by changing a test, without taking mediating factors into account.

A sense of confusion is likely to be reduced if both teachers and students feel that the marking practice is at least perceived to be more consistent. Another way to minimise possible negative washback effects that the assessment is confusing and failing to motivate students is to demonstrate to teachers and students that the assessment is functioning as intended. For teachers, their sense of confusion might be reduced if they feel they have more control over their students' achievement. For "teaching to tests" to be considered positive, both teachers and students need to feel that the role of teachers is to deliver knowledge effectively, and to prepare their students for assessments.

In addition, it would be helpful for test developers to be aware of the conditions for initiating teacher conceptual and behavioural change. In this connection, it would appear important for them to bear in mind that what may possibly affect teachers is not only *what* is included in the test, but also *how* the objectives of the test as well as training is provided to the teachers. As stated above, it would also appear important if test developers are aware that an over-emphasis on the power of tests and absence of attention to teachers' involvement may account for why intended washback did not occur in most studies. To ensure the success of instructional reforms, it is highly suggested that they de-emphasize the power of tests, on the one hand, and encourage teacher engagement, on the other. The goal of analyzing the role of the 'teacher factor' in fostering washback is not to dissuade test designers from using tests as instruments to innovate teaching, but rather to raise their awareness of the important role of this factor in engineering washback so that they may work out better ways to improve the existing test designs.

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

385

On the other hand, it would be beneficial if adequate attention were directed at another one of the stake-holders, language teachers. Given the involvement of the 'teacher factor' in effecting teacher behavioural change, consideration needs to be given to all issues and conditions concerning the observable (i.e., practices) and unobservable dimensions (i.e., beliefs) of teaching. As the chief implementers of reforms, it would appear beneficial for teachers to be aware that innovation and change are a necessary part of teacher development (Bailey 1992; Willis & Willis 1996). It would also appear essential that they be aware of the ways that can be used to promote their awareness. In the meantime, it might be crucial that support be provided to them to promote their awareness of the issues relevant to instructional change.

As mentioned above, one salient finding of the study is related to discrepancies and mismatches in teachers' beliefs and actions in the classroom situation. Based on the theoretical framework developed in the study, the discrepancies may be attributable to the interplay of teachers' Beliefs, Assumptions, Knowledge, and Experience (BAKE) network. More specifically, they may be attributed to the incompatibilities between teacher declarative knowledge and their procedural knowledge (working knowledge). In view of the discrepancies, it would be highly important for researchers to be concerned not only with the declarative dimension of teacher beliefs and knowledge (i.e., understanding of CLT, the test, syllabus and curriculum), but also with their procedural dimension (i.e., the underpinnings of CLT).

The objective of teachers reflecting on Beliefs, Assumptions, Knowledge (BAK), according to Woods (1996), is to facilitate teachers' readiness. Based on his account, an awareness of one's own Beliefs, Assumptions, Knowledge (BAK) may make it easier for one to accept others', to understand how they differ, and to decide that the difference can be worked through in areas of conflict. Other researchers (Borg, 2003; Kennedy 1987, Richards & Lockhart, 1994) also suggest ways to promote teacher awareness. Like Woods, they also deemed reflection as a pre-requisite of teacher development and held that teachers should be encouraged to engage in reflection, research, or systematic inquiry. The benefit of reflective teaching, from their interpretation, is that it may allow teachers to make tacit beliefs and practical knowledge explicit. Here, what should be noted is that underlying the

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

386

view that underscores the important role of reflection is a constructivist view about learning.

Given the myriad of misconceptions about teaching and learning, the limited English proficiency, and the insufficient knowledge and training that Bangladeshi EFL teachers have, it would appear that reflective activities such as training in English teaching methodology, test and curriculum development, research (i.e., observation, introspection and inquiry), and micro-teaching or peer-teaching might be beneficial to them, for such opportunities may allow them to actively reflect on their beliefs, knowledge, and practices, and then modify and reinterpret them. Furthermore, these reflective activities may also allow them to enhance teacher professionalism. The process of reflection may also allow them to reconstruct their personal theories of language teaching and learning, and recount their dilemmas (Crandall, 200; Freeman, 1998; Freeman & Richards, 1996; Shulman, 1992; Woods, 1996). In addition to the reflective activities, the participants in this study may benefit from opportunities to link theory with practice. To be concrete, they may benefit from hands-on guidance that serves to integrate the constructivist theory with the actual classroom practice.

However, neither guidance nor training is sufficient. Teacher conceptual and behavioural change may involve collaborative and/or autonomous learning. The present researcher would suggest that teachers concentrate on improving their methodological skills to achieve effective teaching. It is hoped that this study can help provide teachers with a basis for reflection about language teaching. Since the current teaching methods adopted by EFL teachers in Bangladesh, as shown in the data of this research, do not enable them to accomplish the objectives of the curriculum developed in 2000, it would prove useful for teachers to modify their instructional behaviours to better meet the students' needs.

There seems to be a need for teachers to be sensitised to the potentially bidirectional nature of washback (either positive or negative). In view of the tendency that they are likely to exaggerate the power of test impact, it would be beneficial if they are aware of the nature of washback as well as the intentions underlying the educational innovation. It might also be beneficial if they are aware that there have already been immense changes in external learning, teaching and testing environment in Bangladeshi EFL education. As mentioned above, there

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

387

seems to be a need for teachers to develop their awareness about their own BAKs. It would also be beneficial for them to be aware of the differences in the methodologies that different teachers employ as well as the gap that exists in terms of their perceptions of test impact and its washback effects on them. It is expected that once they see and reflect on the differences and gap in their BAKs as well as in their teaching behaviours, they may be aware of the real problems underlying their teaching practice and probably develop a more reasonable view about their teaching and then be in a position to change their methodologies to lesson the gap.

It would be helpful for teachers to be aware of the urgent need to update their knowledge of language teaching and learning theories, since the traditional transmission model of education still prevails in their classrooms, and language ability is still viewed by some teachers as a set of finite components – grammar, vocabulary, matching, multiple choice pronunciation, etc. Thus, it appears that there is a need that some "input" related to language learning and teaching theories should be provided to them to make them aware that language acquisition concerns more than knowledge of grammar and vocabulary. Above all, their outdated teaching concepts and philosophies might need to be transformed and replaced by more up-to-date ones.

As teachers' awareness of the above issues is increased, hopefully, they would be motivated to reform their own curriculum and make conscious efforts to modify their English teaching methodologies. With a greater awareness and more accurate perception of the nature of washback effects and the conditions under which they operate, teachers may eventually alleviate negative washback and replace it by positive washback. One way of promoting teachers' awareness and facilitating their involvement in the change process, as suggested by Davison (2008), Tavares and Hamp-Lyons (2008) and Turner (2008, 2009) is through providing well-tailored guidance, support and training to them. Such guidance and support may trigger their reflection on some of their internal factors as well as the issues that emerged in this study.

In addition, considering the common resistance encountered in various educational innovations as well as the negative perceptions the EFL teachers maintain towards the HSC examination, it would be helpful if more interaction were undertaken between the interpretative processes of the teacher and test developers. It

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

388

might be helpful for test developers to give a rationale for carrying out a certain type of change. Meanwhile, it also seems to be of help if the objectives and purposes of the HSC exam in English are made available to all test-takers. With an enhanced understanding of the rationale, objectives and purposes of the HSC English, teachers may reframe their conceptions of the testing reforms.

Instructional change occurs as a gradual progress and it is the product of long-term comprehension of different contexts for teaching. To bring about such a change, the EFL teachers may also need to attain adequate proficiency and build a sound knowledge base. For instance, there seems to be a need for them to enhance their language skills and develop a much more fine-grained understanding of the principles involved in CLT or task-based approaches. In addition to guidance on the testing dimension of change, it appears that teachers are badly in need of guidance to correctly interpret the theories of second/foreign language education, and translate them into effective instructional practice. Students seem to have more pathways to promote positive washback and to reduce negative washback than teachers. The system of assessment is likely to be considered helpful for their learning if students feel that external assessments can be trusted and that learning for assessments is encouraging them to think in new ways. To minimise the likelihood of assessment failing to motivate them, students also need to feel that the assessment system is helping them towards their future goals.

The frontloaded curriculum alignment is practiced in Bangladesh English education. That is, the curriculum is developed first. The test is designed to measure how students have learnt based upon the curriculum. One of problems with frontloading alignment is the poor test quality, in terms of lack of validity and reliability. The main goal of the present curriculum of English education is to promote a communicative syllabus in classroom teaching and learning. The HSC EFL test should be written to test students' communicative competence on the basis of the curriculum desire.

Due to its multiple choice, vocabulary and grammar format and excluding oral and aural test, how students' communicative competence can be assessed is questionable. Thus, finding effective ways to include communicative language goals in oral assessment should increase the match of the curriculum and test. Valette suggests if a new test or assessment instrument is seen as particularly valid, its

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

389

availability may exert influence on the statement of desired outcomes and the elaboration of the curriculum" (1994, p. 10). Positive washback is more likely to occur when a curriculum and test are highly matched.

Policy makers are likely to be able to reduce a sense of confusion by modifying the belief that the assessment is interfering with student learning if such a belief is evident among students. It should be recognized, however, that where belief change is required, this is unlikely to be accomplished easily. It will take time and planning. Core values held by some teachers could be threatened by innovation (Fullan, 2001), and they may not feel comfortable even if it was introduced to promote student learning. Thus, teachers should be given sufficient time to learn new ways of teaching. Noble and Smith (1994b) suggested that policy makers should provide teachers with mentoring and a series of professional development courses and/or seminars in order to assist belief change over time.

The current research showed that students' beliefs were influenced by negative images of the new assessment system painted by the media and adults around them. Policy makers would need to allow students and the whole community to adjust their beliefs, when necessary, with multiple opportunities over a prolonged period through publicity and education to learn about the rationale for a new assessment system and its benefits. Such strategies could go a long way to fostering the development of positive beliefs about the new assessment, in turn fostering positive washback.

The present research aimed to build a much needed bridge between the fields of education and applied linguistics in washback research by bringing theories of both fields into one study. The proposed model has demonstrated part of the mechanism of washback, while it has confirmed its complexity as discussed in the relevant literature. Assessment will probably continue to be used around the world as a tool to bring about improvement in teaching and learning. Further understanding of the mechanism of washback can help reduce undesirable negative washback and can help utilise the full potential of the power of assessment to bring about desired educational change. It is hoped that the current investigation might stimulate dialogue among researchers with the aim of contributing to the understanding of the mechanism of washback. It is expected that this study may bring a noticeable changes in the field of English language teaching and learning at the HSC level if

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

390

the suggestions put forward in this study are brought into consideration for implementation.

The above situation suggests that tests might not be a good lever for change automatically if this lever is used to set the educational machinery in motion. The machinery, in the sense of the educational systems or teaching practices, will not easily lend itself to control by tests. Under such circumstances, the challenge confronting the testing experts in Bangladesh is to work out a sound testing policy based on rigorous validation research, which will help maximise the effectiveness of English testing for learners. It is hoped that language testing research can ensure that this powerful instrument does not victimise the innocent, but serves its purpose by enhancing the prosperity of the public and society.

6.4 Recommendations

The results of the study suggest that improvements should be made in future reform efforts in terms of test design, teacher training, innovation of the syllabus and curriculum, and research methods. Some recommendations are made with the intent to facilitate the implementation of educational reforms or innovations for promoting positive washback. This section outlines a set of recommendations in relation to the findings of this study.

6.4.1 Recommendations for Improving the HSC Examination in English

From the present study, it is now a grounded fact that the present HSC Examination in English suffers from lack of validity and reliability. The study finds that the HSC examination is a problematic one. Tests objectives need to be clear and transparent. As a means of promoting teacher pedagogical change, they need to reflect a shift from transmission, product-oriented theories to constructivist, process-oriented theories and pedagogies of teaching and learning. Teachers can be in a better position to change their perceptions and behaviours (which conform to its innovation) only when they achieve a good comprehension of the objectives, content and methodology of a test. It does not correspond to the aims and objectives of the

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

391

syllabus and curriculum. The contents and test items of both *First and Second papers* cannot test adequately the competence or proficiency in English.

Examinations should reflect the syllabus and curriculum, and since not everything in a curriculum can be tested in an examination, the areas that are assessed should be ones that are considered important. It is also important that, the same items and contents should not be tested again and again. If the expectation is that student achievement will align itself with the content of examinations, it is critically important that the stated objectives and content be carefully developed. Test objectives should reflect the contribution of knowledge and skill, which they embody to the long-term growth and development of students. On the bases of findings, the researcher would like to put forward some recommendations so that it can promote a beneficial washback on the EFL teaching and learning at the higher secondary level in Bangladesh, and thus, the objectives of the curriculum be achieved. The recommendations are stated below:

1. Provision of testing of listening and speaking skills should be organised in any acceptable format such as TOEFL, IELTS, MELAB, TSE, etc.

2. Insofar as possible, modes of testing (e.g. written, practical, oral) should be diverse to reflect the goals of curricula.

3. The EFL examination items should not be limited to the measurement of recall or recognition of information, but should attempt to measure higher-order outcomes defined in terms of more complex cognitive processes (e.g. understanding, synthesis, application).

4. The EFL examination should assess students' ability to apply their knowledge, not just in scholastic contexts, but also in situations outside school.

5. Examinations, both in content and in difficulty level, should reflect their certification function and provide an adequate assessment of the competencies of pupils who will not obtain a higher-level qualification.

6. The EFL examination performance should be analysed to provide feedback to schools and other stakeholders (e.g. curriculum authorities, etc.).

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

392

7. As many teachers as possible should be involved in the setting and scoring of examinations, as these activities provide valuable insights into the demands of examinations which can be applied in teaching.

8. Items such as of questions 11, 12, and 13 in the *First Paper* should be revised and reconsidered because those items, carrying 40 marks, are over credited. Besides, questions 11 and 12 can hardly test any linguistic competence.

9. The question papers (*First Paper & Second Paper*) do not maintain the difficulty level. So, difficulty level of the HSC examination items should be maintained.

10. Repetitions of items should be minimised.

11. The use of teachers' assessment to contribute to grades that their students are awarded in public examinations merits investigation.

12. Testers/question setters and examiners should be trained properly so that they can perform their respective duties in view of communicative language teaching and testing at the HSC level in Bangladesh.

13. Guidelines for EFL examiners should be formulated for examining and scoring the answer scripts.

6.4.2 Recommendations for Curriculum and Textbook Revision

The HSC curriculum was formulated nearly a decade ago; it was then well-suited for teaching and learning English. But, revision of the curriculum is a continuous process on the basis of needs and demands of students and time. The findings of the study revealed a mismatch of intention and execution between policymakers and implementers. Teachers as implementers did not carry out the intended curriculum. Rather, they conducted classroom instruction based on the context and reality where they were teaching.

An obvious practice of teaching to the test is often referred to as a narrowing of curriculum and instruction. Narrowing of the curriculum and instruction is not easily identified as appropriate or inappropriate. For example, a positive effect is that it guides course of study revisions and lesson planning by emphasizing certain

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

393

areas of the curriculum or, in other words, it helps teachers sharpen their focus. However, a negative effect is that it decreases the breadth or depth of content and activities to which students are exposed because the test does not emphasize the content or skills that some activities address:

1. The syllabus contents and form of examination items should exhibit correlations with the EFL testing.

2. The mother textbook *English for Today* for classes 11-12 was written in 2001. So, it is high time the National Curriculum and Textbook Board (NCTB) authority revised the textbook introducing more current authentic materials, lessons and exercises.

3. The use of commercially produced clone test materials (hidden syllabus) in the class should be discouraged.

4. The syllabus and curriculum contents should be modified and reshuffled after every three years of time because learning needs and context change very rapidly as per the demand of the society.

5. The authority concerned should monitor the teaching and learning activities in the classroom, testing, and check the test related materials whether they enhance EFL learning.

6. The present HSC English curriculum is considered to be a frontloaded one. The whole syllabus of the English curriculum is accommodated in the textbook (*English for Today* for Classes 11-12). Some of the lessons are never tested since it has been formulated; even those lessons are never taught by the teachers because those are complex and thematically uninteresting. Those lessons should be replaced immediately.

7. Testing is an integral part of any curriculum. All formal syllabuses make provision of assessing how much of the syllabus is taught; how much the learners have learned, and how much the curriculum objectives are achieved.

8. The effectiveness of commercial test-preparation materials used for preparation should be monitored or banned so that teachers and learners can be compelled to use the prescribed textbook.

9. The teachers are reluctant to teach the syllabus in the class due to test pressures. They teach the selected topics which are likely to be tested in the examination. The findings of this explanatory case study revealed that there are not so many inconsistencies between the learners' needs, and the textbooks that are available for learning and teaching English language. Therefore, the authority should ensure that the teachers teach the textbook properly.

6.4.3 Teacher Training for Promoting Beneficial Washback

The role of teachers as one of the main stakeholders in the implementation of educational change and curriculum innovation has been the focus of ongoing interest to curriculum researchers, and has been examined extensively in the literature, both conceptually and empirically. This is because teachers, as implementers, determine whether or not curriculum innovation is executed in the classroom as it is intended by policymakers. Teachers are the main implementer of the syllabus and curriculum. McLaughlin (1987) suggested that at the institutional level, teacher training should be offered, and human, financial, and material support should be provided to teachers to facilitate their implementation.

For instance, Carless (1999a) asserts that teachers are the individuals who implement, adapt, reject, or ignore curriculum innovation. It is thus something of a truism that they are the core of the innovation process. Therefore, teachers' decisive roles in the implementation of the curriculum cannot be underestimated. Without their support and genuine involvement in the innovation, any curriculum implementation will stay at a superficial level, with either semi- or even non-implementation. Therefore, the following recommendations are made for their improvement:

1. Future testing innovation endeavours need to be accompanied by ongoing training and/or appropriate teacher guidance and support on assessment and instructional practices over time. As teacher evolution is a transformative process, it takes time for cognitive and behavioural changes to take place. Thus, teacher training for language teachers should provide adequate opportunities to engage them in reflective activities (e.g., classroom research,

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

395

action research, classroom observation, team teaching, test and curriculum development), since these activities, as noted by Richards (2008), are the principal sources for constructing a knowledge base.

2. Teacher training needs to help teachers raise their awareness of what their pedagogical knowledge is and whether there is a gap between their pedagogical knowledge and that defined by FL theorists. As discussed above, this is considered to be a crucial step for developing teacher expertise. From Woods' (1996) point of view, teachers' beliefs, knowledge and practice evolve through awareness of these discrepancies and resolution of conflicts. He puts forward the suggestion that teachers be provided with more opportunities to reflect on their teaching and recount their dilemmas. According to him, it is the dilemmas that cause teachers to analyze and reflect on their beliefs, and to consider the various options for achieving their teaching goals. In this regard, rather than viewing conflicts, differences and discrepancies as forces that hinder the processes of change, they should be perceived as an impetus for change and an indispensable part of teacher evolution.

3. Teacher-training programmes need to help teachers increase their awareness of various discrepancies that may exist in their perceptions and behaviours. Meanwhile, guidance needs to be provided to teachers, as suggested by Richards (2008) and Woods (1996), to make their tacit and inaccessible beliefs and knowledge of teaching pedagogy explicit and accessible.

4. Research on washback needs to focus on the longer-term effect of teacher change (i.e., change over time). Owing to the nature of the nested 'teacher factor', research needs to examine the phenomenon of washback in the growing professionalism of the field. Consideration should be given to how teachers' beliefs (cognition underpinning teacher practice) are conceived, how their knowledge structures are formed, and how teachers' beliefs and knowledge inform their practice.

5. The notion of washback needs to be redefined or re-conceptualized. Rather than dwelling on debates over whether a test can induce positive or negative washback, an emphasis should be given to the more complex issue of how a

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

396

test interacts with the 'teacher factor' in the course of washback. To deemphasize the power of tests, consideration needs to be given to how to secure teacher consent, participation and active involvement in the change process. As discussed in Section 6.2.3 above, apart from the studies conducted by Davison (2008), Muñoz and Álvarez (2010), Tavares and Hamp-Lyons (2008) and Turner (2001, 2002, 2005, 2008, 2009), it appears little attention has been paid to teachers' active involvement in innovation. However, 'teacher involvement' seems to be a factor that should not be overlooked in washback research (Turner, 2008, 2009). Based on Woods (1996), such an over-emphasis on the power of tests and absence of attention to teachers' engagement may account for why intended washback did not occur in most studies. Therefore, to ensure the success of instructional reforms, there is a need to de-emphasize the power of tests on the one hand and encourage teacher participation on the other. Teacher training provides teachers with innovation of teaching, builds awareness of responsibility, guides how to shape their teaching for the best outcomes, informs of power of tests, etc. Training will persuade teachers to pay serious attention to communicative skills and to teaching learning activities that are more likely to be helpful in the development of such skills. Therefore, teacher training can be considered a powerful tool for generating beneficial washback of examination.

6. Research from general education needs to be drawn on in research on washback. It appears that studies concerning the role of the 'teacher factor' in the context of language testing are still scant. The rationale for drawing on other areas of research is that there is a substantial body of literature in these areas that focus on beliefs, knowledge, proficiency, experience, and practice. The insights gained from these areas may help us tease apart the 'teacher factor' in washback.

7. Efforts need to be exerted to encourage classroom-based performance assessment (e.g. role plays, interviews) and personal-response assessments (e.g. self-assessments, conferences) (Brown & Hudson, 1998), for these assessments are considered to be a bottom-up process. The application of this

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

397

bottom-up, process-oriented assessment approaches may help balance those which are top-down, product-oriented (Colby- Kelly & Turner, 2007).

8. Great importance must be attached to the understanding of the socio-cultural, educational and institutional context in which teaching, learning and testing occur when planning and implementing reforms.

9. Greater emphasis should be placed on alignment of assessment and the central agents of change since it appears that considerations and efforts directed at alignment of assessment, curriculum, teaching and learning practices (as Colby-Kelly and Turner (2007) calls it, 'assessment bridge') do make a difference.

6.5 A Washback Model Proposed by the Researcher

The research investigated the washback of the HSC examination, especially on the teaching and learning English as a foreign language. High-stakes tests have been commonly used in many countries in an attempt to bring about intended positive washback on teaching and learning. The small number of washback studies in applied linguistics to date has led to a limited understanding of washback effects on language teaching and learning, and to a lack of a widely accepted model which illustrates the mechanism of washback in the field. The lack of such a model also applies to the field of education despite decades of test impact research. In going beyond simply identifying the nature of washback, the major contributions of this research have been to help further understand how intended positive washback effects are promoted or hindered by certain contextual factors and beliefs held by teachers and students, and to propose a washback model which demonstrates the role of these factors and beliefs in mediating washback.

Furthermore, the assessment does not seem to affect teachers and students in a uniform manner. This appears to illustrate the complexity of the process of test impact. Numerous factors are likely to be involved in such a process. These are reflected in the recent models of washback proposed in the field of applied linguistics. Such models still do not articulate, however, exactly how various factors play a role in mediating certain types of washback effects. In this thesis, the argument is presented that intended washback effects are more likely to be brought

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

398

about, or at least washback may become more predictable, if the link between mediating factors (external and internal) and washback is made more apparent.

Based on the proposed washback models discussed in Chapter 2.5 (sections 2.5.1 to 2.5.1.1.3) and the findings of the present study, the present researcher proposes a model of the washback of the HSC examination on EFL teaching and learning. A test is a part and parcel of a curriculum, but it is an activity outside of the classroom. A number of factors and a group of participants (Hughes, 1993) are directly involved in teaching, learning and testing process. The model below (Figure 6.1) shows how positive washback at multi-level may be generated:

Figure 6.1: A washback model proposed by the researcher
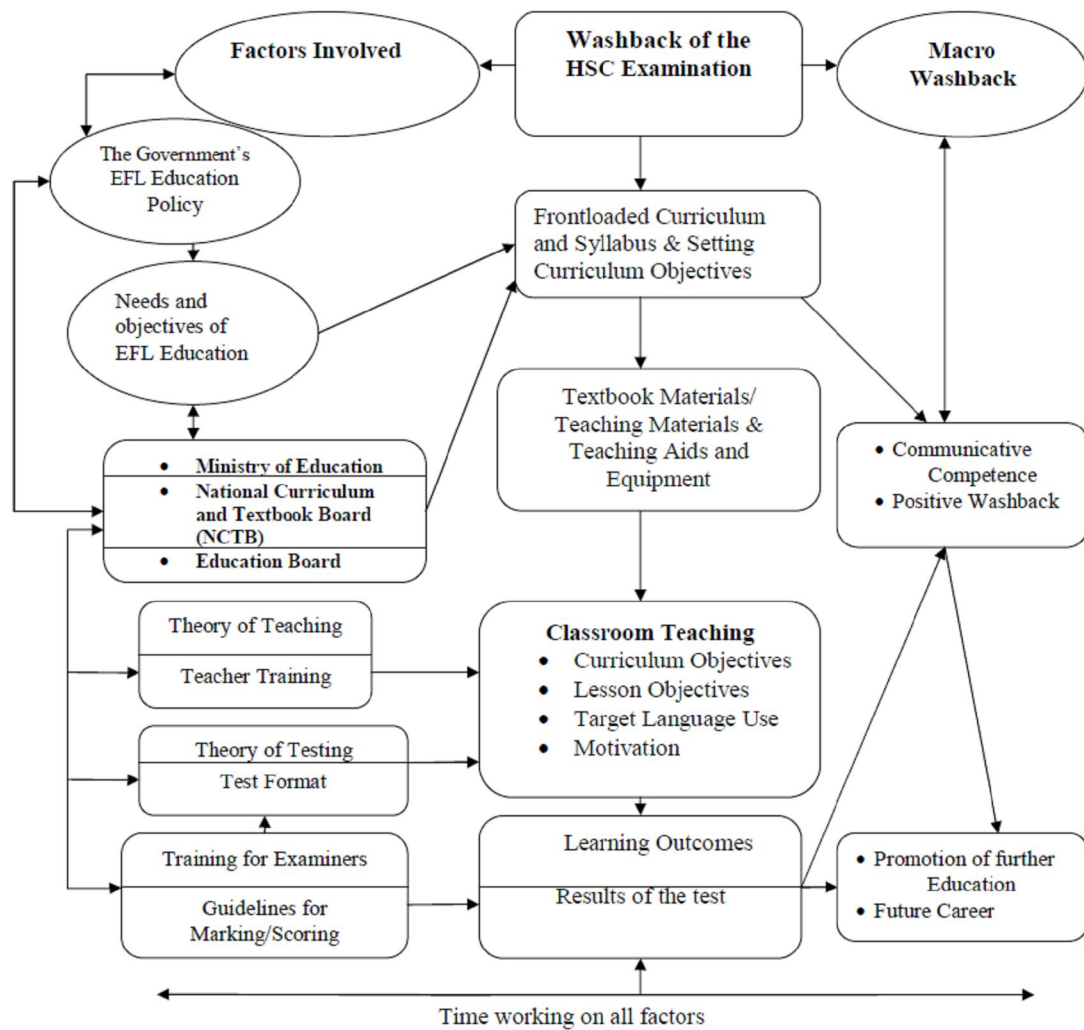
The present researcher proposes the model generated from the previous analysis of washback studies, the major washback models, and current leading

theories such as Alderson and Wall's (1993) fifteen washback hypotheses, Bailey's basic model of washback (1996), Pan (2008), Shih (2007,2009), Green (2003), Tsagari (2009), Cheng (1999, 2002,) and Hughes' trichotomy of washback (1993). In comparison to Bailey (1996) and Cheng (2002) model, which lays out factors that contribute to washback along three specific categories(participants, process, product), this model demonstrates not only that participant, process and product factors contribute to washback – the sub-elements within each factor may interact among themselves or across factors to create a washback effect.

As shown in the model, a macro level positive washback requires a number of direct actions that are related to the classroom activities and testing. The government should adopt EFL education policy and set needs and objectives of language education. To fulfill the EFL education objectives, a frontloaded curriculum is a first and foremost requirement because a curriculum is a vital part of EFL classes. The syllabus and curriculum should set a number of objectives that should be achieved through classroom teaching. In accordance with the objectives of the syllabus and curriculum, textbooks should be written so that it can correspond to the objectives of the syllabus and curriculum. To promote positive washback, classroom tasks and activities should be based on the textbook materials. The arrows in the model (Figure 6.1) indicate how the different factors are interrelated.

The model shows that the teachers and examiners of English should be trained up so that they can effectively contribute to teaching and learning in line with the curriculum objectives. Test format should be based on the classroom practice. The model shows that testing is a subordinate to classroom teaching, and the servant of the syllabus and curriculum. The model shows that the outcome of teaching is communicative competence which is tested in the examination. Time is an important factor for producing teaching and testing results. Shohamy, Donitsa-Schmidt and Ferman (1996) have pointed out that washback evolves with the passage of time. The product of learning outcomes and test results promote further education and future career.

The model above (Figure 6.1) shows that three entities (ministry of education, NCTB, and Education board) of the government are directly responsible for EFL education in Bangladesh. The model indicates that NCTB is designated to formulate the curriculum and syllabus. The syllabus and curriculum should include a

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

400

wide range of topics from both national and global contexts. It provides a focus for the class and sets goals for the students throughout their study. A curriculum also gives students a guide and idea to what they will learn, and how they have progressed when the course is over.

The model corroborates what Shih (2007, p.137) has mentioned that "any given test needs research tailor-made to examine its washback". It also expands upon the model proposed by Shih (2007), which examined how multiple factors may variably impinge on washback of student learning only. In contrast, this model shows very distinctly how very specific factors, and combinations of factors drive the washback effect on both teaching and student learning: what teachers decide to implement in their classrooms relative to a high-stakes exam inevitably impacts how students respond to it, meaning that the two are inextricably linked.

The results of this study suggest that in educational systems such as the one present in Bangladesh, looking at only one or the other presents an incomplete profile of the washback effect of any given high-stakes test or examination. It also adds a dimension that was missing in the previous washback models suggested by Bailey (1996), Cheng (2002), Pan (2008), Saif (2006), and Shih (2007) – the dimension of time. The model that the present researcher is proposing is grounded in empirical evidence showing the effect of time on washback at two levels. The model suggests that examining the washback effect at different periods of implementation shows a different washback profile in terms of teaching and learning. Details of the procedures of positive washback are reflected in the proposed model. It is expected that HSC examination may promote macro-level positive washback if the above model is properly followed and implemented.

6.6 Suggestions for Future Research

This research has extended the understanding of the mechanism of washback. Future research, however, could expand on certain aspects. Further research is needed of "a phenomenon on whose importance all seem to be agreed, but whose nature and presence have been little studied" (Alderson and Wall, 1993, p. 115). The study has highlighted a number of areas where further research is needed. The evidence that examination preparation courses change over time

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

401

supports the inclusion of a longitudinal dimension into future washback studies. Alderson 1998 summarises the need for future research:

> In an ideal world …the way teachers prepare students for examinations would look no different from how they teach every day, and how they teach every day would match both the aims and the methods that the examinations intend to inculcate, and the methodology that the textbooks and the teacher education courses advocate. However, it is absolutely clear that teachers will indeed engage in special test preparation practices, and therefore it is important to consider what best practice in test preparation should look like.

Considering the complex nature of washback, and in view of what was found in this study (that exams alone have a limited impact on how teachers teach), the present researcher is now in a position to make a few suggestions for future research. He suggests a shift of focus from discussions of the existence or nature of washback in the field to the study of the role of the internal factors (e.g. the 'teacher factor') as well as of the external factors (e.g. the 'learner factor') in the process of washback. Intensive research should also be carried out on the washback effect on the syllabus and curriculum, teaching materials, teaching method, etc. Another avenue that has not been explored in the present washback research is the use of student test scores to measure washback. Although the present study conducted the test analysis as on of instruments of the research, it requires more exclusive and thorough study involving question paper setter and more examiners. It would be challenging to measure the actual extent of washback by simply calculating the test means for students.

In addition to asking whether washback exists or not, and whether it is negative or positive, it should first be asked how a test can be developed to contribute to fundamental teacher methodological changes. Future studies could integrate teacher beliefs about personal constraints within the context of language instructional innovation, giving teachers opportunities to discuss and reflect not only about their beliefs about teaching, learning and how to teach, but also about personal constraints, dilemmas and challenges encountered in carrying out instructional practice. In order to understand the teacher washback behaviours more thoroughly within foreign language classes, studies need to be conducted in classrooms at different levels.

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

402

According to Richards and Farrell (2005) and Woods (1996), stimulated recall would give a clearer picture of what can trigger reflection. They view it as an appropriate method for revealing the underlying beliefs or motives of teachers. According to them, this method may allow teachers to describe or articulate what they know about teaching and learning and how they know it. Thus, it would be beneficial for future research to conduct studies using this method to further examine the values and beliefs underpinning teachers' classroom practice.

As mentioned in Chapter Five, this study has chosen to examine some aspects of the 'teacher factors' (i.e., teacher BKEs). In this regard, it would be imperative and informative if other aspects are integrated into future research. One suggestion is that future research can broaden the scope to include aspects such as teacher proficiency, teacher background, teacher motivation, etc. In order to understand the teacher washback behaviours more thoroughly within foreign language classes, studies need to be conducted in classrooms at different language levels. Conducting the present study in an intermediate language classroom at the secondary level to determine if there are notable differences between the two levels could increase the generalizability of the present study.

The results of the study have demonstrated that there are several other factors which need exploring. Therefore, more extensive studies need to be conducted to explore how various forces are combined to produce varying cases of washback. In addition, future research needs to further examine the 'student' variable in washback. In fact, this is a concern commonly shared by other researchers (e.g. Alderson, 2004; Bailey, 1996; Messick, 1996; Muñoz & Álvarez, 2010). They all hold that for beneficial washback to take place, the student role and/or student awareness of examination objectives and their connection to educational goals must be taken into consideration.

When considering other studies which focus on the perspective of the teacher, such as Watanabe (1996) and Alderson and Hamp-Lyons (1996), the benefit of observing teachers teaching both a test preparation and a non-test focused course with students at a similar language level is significant. This would help identify whether the methodology a teacher employs is particular to teaching test preparation or is also the approach they apply to other teaching situations.

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

403

Future research should gather evidence from different populations and high-stakes tests so that the phenomenon of washback of public examinations can be understood more thoroughly. Some studies have suggested that student factors contribute to explaining how washback of public examinations influences how students learn and how different types of learning attitudes might indirectly influence the way teachers teach (Cheng, 1999). Other studies pointed out those studies of washback effect on different language tests, across settings, across cultures, and across language backgrounds, increase future generalizability regarding washback effect of public examinations on language education (Saif, 1999). Replications may provide greater generalizability regarding washback of public examinations.

A high proportion of the interviewed teachers indicated that they were concerned about their students' abilities when developing their lesson plans. Another noticeable proportion the interviewed teachers revealed that the constraint of time and Education Board policy had an influence on their curricular planning and instruction. Therefore, other variables, including students' abilities, time and policy should be included in future studies.

To the best of the present researcher's knowledge, this research was the first to demonstrate exactly how washback was mediated by a number of teaching learning factors and beliefs. Obviously, further investigation is necessary to determine whether the patterns identified can be generalised to other settings. The current research also found that there were more pathways to positive and negative washback effects for students than for teachers. That is, there were more opportunities to promote positive washback, while there were also more possibilities for negative washback where students were concerned. This was an interesting finding which merits further investigation.

While the influence of beliefs in washback was identified in the current research, one might argue that the influence could also work the other way (i.e., teachers' and students' experience with assessment could affect their beliefs). This line of inquiry was beyond the scope of this research, but in reality, it would be very likely to see interactions between beliefs and washback. It would be a chicken-or-egg question, but worth exploring; the future researchers should forward to unfold the real truth that may generate positive washback on the EFL teaching and learning

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

404

in Bangladesh in general and at the HSC level in particular. Although the present research established a link between washback and students' expected (i.e., perceived) achievement, investigating a link between washback and students' actual learning outcome was also beyond the scope of this research. It would be useful for future research to investigate the relationship between the two, given that educational reform through standards-based assessment tends to aim explicitly to improve the learning of students of all abilities.

6.7 Conclusion

Washback is a challenging phenomenon to research and measure. Given the variables present in the classroom, this challenge persists in washback research. Although this study focused primarily on teaching and learning, teacher behaviours in the classroom, the investigation into the students' perceptions of washback and the students' behaviour demonstrating washback, it covered some other domains (e.g. the syllabus and curriculum, materials, teaching methods, etc) which are affected by washback. Every teacher wants his/her students to perform well on tests as it is a reflection in his/her teaching. This causes negative washback

The important factors influencing changes in teachers' behaviours reflect teachers' beliefs, and the students' desire to have accurate information about the testing programme. Teacher beliefs influence teacher behaviours deeply no matter how the curriculum, course objectives, and the testing programme work together. Students want information about the testing programme because they want to be well prepared for the test as a result of the classroom instruction. This study supports the notion that in order to affect the instruction that occurs in the classroom and create positive washback, direct guidance and training must be part of a teacher's professional development. Even experienced teachers need professional development training in the areas that are new to them when teaching a course for the first time. As washback studies by Cheng, (2002), Watanabe (2004), and Shohamy et al. (1996) have indicated, tests often provide teachers with new reasons for new activities; however, they do not necessarily show the teachers how to incorporate these activities into the curriculum.

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

405

Additionally, the tenets of past traditional language learning paradigms and their strong influence on present teacher behaviours as suggested in this study must not be ignored. Whether the teachers in this study were once students in a foreign language programme that was heavily influenced by grammar, vocabulary, multiple choice questions, matching, and structure, or whether they were trained or worked in a programme that influenced these areas in a previous teaching assignment, grammar appear to strongly influence the teachers' current instructional practices.

This study shows that washback behaviours focusing on grammar, vocabulary, matching/rearranging, and structure are evident in instruction, regardless of the design of a language programme, the format of the courses and the materials, and the demands of the testing programme. Watanabe indicated similar teacher behaviours in his washback study (2004b), as it seems that teachers finds a place in their repertoire for specific and direct grammar instruction. Focusing attention on this aspect, especially for first time teachers in an intermediate level, skill focused language class, can assist controlling authority as they help teachers to be more effective. More efficient teachers mean happier and more successful students.

Tests are the part and parcel of any curriculum. Tests can never be eliminated from educational institutions; therefore, it is best to embrace them and their power. Only then might they be used as a tool to bridge the gaps in instruction, and to train teachers not only to meet the demands of tests, but also meet the demands of their students, and affect their students' achievement. Tests are only one component affecting change in the classroom; guidance and professional development are also necessary to the process. With focused guidance and professional development, this study implies that progress can be made toward creating positive washback from tests in the classroom both in terms of teacher behaviours and student behaviours. Tests can drive change, and if the intention is to make changes to foreign language instruction, not only at the intermediate level, but at all levels, it must be done consciously. Efforts must be taken to help teachers encourage positive washback and reduce negative washback.

In higher secondary education (the context of this study), language education programmes must focus on preparing teachers to teach well and teach effectively. Teachers at this level may need adequate training to balance their limited experience, and a testing programme can provide the springboard to this training.

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

406

Additionally, teachers also exhibit positive washback behaviours, and thus set the example for the students that skills are the driving force in the classroom, as they happen to complement the tests.

With respect to the literature on washback, the inter-disciplinary, mixed methods approach used in this study, which responds to calls from previous research conducted in this domain. It demonstrates that, indeed, such an approach may provide an interesting avenue for testing apart the complex interactions and effects involved in washback. A positive consequence of washback behaviours that reflect skills can be successful test performance. The findings of the study contribute to knowledge about an as yet unexplored context in the washback literature. It demonstrates the uniqueness of interactions between factors within each educational context in creating specific washback effects: the model of washback proposed may inform research designs of washback in novel contexts.

As indicated at the beginning of this paper, the HSC examination in English has been widely accused of its "adverse effects" (so-called "negative washback") on teaching and learning teaching and learning. When the results of this study are compared to research carried out in other EFL contexts, two main characteristics are found to exist in current testing and educational innovations. One characteristic relates to the way EFL teachers in Bangladesh perceive and react to testing and ELT innovation. This is similar to that exhibited in other contexts. Such a characteristic reflects the fact that while tests in different contexts may affect teaching and learning at varying degrees, and moreover washback may take different forms, the rules or mechanisms underlying the washback phenomenon as well as the complexities implicit in it seem to make little difference.

Currently, new technologies and changing foreign language education environments provide meaningful and novel ways to incorporate communicative language tests, a real challenge in the past, into EFL curricula. In fact, this phenomenon draws increasing attention with contemporary changing perspectives of the definition of literacy: from the conventional reading and writing focus to the more functional and current electronic literacy for students in a future global environment. In line with these shifting trends, the results of this study highlighted some important issues and posed several important recommendations for future research on English competence assessment in EFL contexts. This study found that

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

407

certain characteristics of test takers might produce a significant influence on test results during an oral English test. Therefore, this research recommends that a multifaceted effort be made to ensure that a test functions fairly across various test takers, regardless of their individual backgrounds.

This study also suggests sharing ownership of testing among test makers, test takers, test users, and examiners which allows all of interested parties to have an opportunity to benefit appropriately from a test. For example, test users collect psychometric information in accordance with the purposes of test use such as selecting or placing candidates. Testers use the psychometric information to develop better tests. Test takers also use psychometric information to improve their ability of interest. Thus, as testing is an interrectual property of various stakeholders, profit motives for testing should not be conflict each other. The study has corroborated previous research and studies that call for a focus on the appropriate use of EFL test formats according to the purposes of the tests and the characteristics of the individuals who take certain speaking tests in English.

The findings of this study have allowed various insights into test preparation classrooms and the HSC examination preparation classes in particular. The researcher hopes that future studies on washback will be carried out to further add to the current picture of the power of language tests to influence those most intimately affected by them. In fine, it is hoped that the recurring features implicit in the change process discussed in this study will provide the basis for improvement for further innovation endeavour. It is also expected that this study will give a strong impetus to the study of the potential areas in washback in future to help generate beneficial washback on teaching and learning English as a foreign language.

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

408

References

Abraha, S., A. Beyene, T. Dubale, B. Fuller, S. Holloway, & E. King. (1991). What Factors Shape Girls' School Performance? Evidence from Ethiopia. *International Journal of Educational Development*, *11*(1), 107-118.

Agresti, A. (1983). A survey of strategies for modeling cross-classifications having ordinal Variables. *Journal of the American Statistical Association*, *78*, 184-198.

Aldridge, A. and K. Levine. (2001). *Surveying the Social World: Principles and Practices in Survey Research*. Maidenhead: Open University Press.

Alderson, J. C. (1986).Innovations in Language Testing? In Portal, M. (Ed.*),* *Innovations in Language Testing: Proceedings of the IUS/NFER Conference*. Philadelphia: NFER-Nelson

Alderson, J. C. (1987). *Innovation in language testing. Can the micro-computer help?* Special report number 1, Language Testing Update. Lancaster: University of Lancaster.

Alderson, J. C. (1990). Learner-centred testing through computers: Institutional issues in individual assessment. In J. H. A. L de Jong & D. K. Stevenson, (Eds.), *Individualizing the assessment of language abilities*. Clevedon: Multilingual Matters.

Alderson, J. C. (2004). Foreword. In L. Cheng, Y. Watanabe, & A. Curtis, (Eds.) *Washback in language testing: Research contexts and methods*. (pp. ix-xii). Mahwah, NJ: Lawrence Erlbaum Associates.

Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics, 14*(2), 115-129.

Alderson, J. C. & Wall, D.(1993). Examining Washback: the Sri Lankan Impact Study. *Language Testing, 10*, 41-69.

Alderson, J. C., Clapham, C., & Wall, D. (1995). *Language test construction and evaluation. Cambridge*: Cambridge University Press.

Alderson, J. C., & Hamp-Lyons, L. (1996). TOEFL preparation courses: A study of washback. *Language Testing, 13*(3), 280-297.

Alderson, J. C., & Wall, D. (1996). *Editorial. Language Testing 13*(3), 239-240.

Alderson, J. Charles and Banerjee, Jayanti (2001) Language testing and assessment : state of the art review, part one. *Language Teaching*, *34* (4), 213-236.

Al-Jamal. D. and Ghadi, N. (2008). English Language General Secondary

Certificate Examination Washback in Jordan. *Asian EFL Journal*, *10*(3), Retrieved September 21, 2009 from http://www.asian-efl-journal.com.

Allen, P., M. Frohlich and N. Spada. 1984. The communicative orientation of language teaching: An observation scheme. *TESOL '83: The Question of Control*. Washington, DC: TESOL.

Allwright, D. & Bailey, K. (1991). *Focus on the language classroom.* New York: Cambridge University Press.

American Language Academy.(1997). [Advertising flyer]. *TOEFL Mastery: Computer-based TOEFL test preparation.* Rockville, MD: American Language Academy, ESL Software Department.

Anderson, J. (1993). Is a Communicative Approach Practical for Teaching in China? Pros and Cons. *System, 21*(4), 471-480.

Andrews, S. (1994). Washback or washout? The relationship between examination reform and curriculum innovation. In D. Nunan, R. Berry, & V. Berry (Eds.), *Bringing about change in language education: Proceedings of the International Language in Education Conference 1994* (pp. 67-81). Hong Kong: University of Hong Kong.

Andrews, S. (1994b). The washback effect of examinations: Its impact upon curriculum innovation in English language teaching. *Curriculum Forum, 1,* 44-58.

Andrews, S. (2004). Washback and curriculum innovation. In L. Cheng., Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp.37-50). Mahwah, NJ: Lawrence Erlbaum Associates.

Andrews, S., & Fullilove, J. (1994). Assessing spoken English in public examinations-why and how? In J. Boyle & P. Falvey (Eds.), *English language testing in Hong Kong* (pp. 57-86). Hong Kong: Chinese University Press.

Andrews, J., Majer, J., Sargeant, D., & West, R. (2000). Reforming language examinations as classroom research: Washback and wash forward in a cluster of teacher training colleges in Poland. In M. Beaumont & T. O'Brien (Eds.),

Andrews, S., Fullilove, J., & Wong, Y. (2002). Targeting washback: A case study. *System, 30,* 207-223.

Allwright, D. (1988). *Observation in the language classroom*. London: Longman.

Allwright, D., & Bailey, K. M. (1991). *Focus on the language classroom: An introduction to classroom research for language teachers*. New York: Cambridge University Press.

410

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

Anderson, M. B. (1988). *Improving Access to Schooling in the Third World: An Overview. BRIDGES Research Report Series, 1,* Cambridge MA: Harvard University.

Appleton, S. (1993). *Gender Differences in Educational Attainment in Kenya: Exam Performance and Access to State Secondary Schools*. Unpublished. McNamara Fellowships Program. Final report. Washington, D.C.: World Bank.

Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.

Bachman, L. & Cohen, A. D. (1998). Language testing - SLA interfaces: An update. In Bachman, L. & Cohen, A. (Eds.), *Interfaces between second language acquisition and language testing research*, pp 1-31. Cambridge: CUP.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University Press.

Bachman, L. F. (2000). Modern language testing at the turn of the century: assuring that what we count counts. *Language Testing, 17*(1), 1-42.

Bachman, L. F. & Palmer, A. S. (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.

Bailey, K. M. (1996b). The best laid plans: Teachers' in-class decisions to depart from their lesson plans. In K. M. Bailey & D. Nunan (Eds.), *Voices from the language classroom: Qualitative research in second language education* (pp. 15-40). Cambridge: Cambridge University Press.

Bailey, K. M. (1998). Learning about language assessment: Dilemmas, decisions, and directions. Pacific Grove, CA: Heinle & Heinle.

Bailey, K. M. (1999). Washback in language testing. TOEFL Monograph Series, Ms. 15. Princeton, NJ: Educational Testing Service.

Bailey, K. M., Bergthold, B., Braunstein, B., Jagodzinski Fleishman, N., Holbrook, M. P., Tuman, J., Waissbluth, X., & Zambo, L. (1996). The language learner's autobiography: Examining "apprenticeship of observation." In D. Freeman & J. Richards (Eds.), *Teacher learning in language teaching*, pp. 11-29. Cambridge, UK: Cambridge University Press.

411

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

Barcelos, A. M. F. (2003). Researching beliefs about SLA: A critical review. In P. Kalaja and A. M. F. Barcelos (Eds.), *Beliefs about SLA: New research approaches*, 7-33. Dordrecht: Kluwer Academic Publishers.

Barnes, M. (2010). *The washback of the Test of English as a Foreign Language Internet-based test (TOEFL iBT) on English language programs in Vietnam*. Ph.D. Thesis. Retrieved December 17, 2010 from www.edfac.unimelb.edu.au/

Baumgart, N., & M. O'Donoghue. (1989). *Improving the Quality of Education in Primary and Secondary schools: A Draft Plan to Improve the System of Assessment and Examinations in Mauritian Schools*. Canberra ACT: International Development Program of Australian Universities and Colleges.

Beattie, M. (1995). New prospects for teacher education: Narrative ways of knowing teaching and teacher learning. *Educational Research, 37(1),* 53-70.

Berry, V. (1994). Current assessment issues and practices in Hong Kong: A preview. In D. Nunan, R. Berry, & V. Berry (Eds.), *Bringing about change in language education: Proceedings of the International Language in Education Conference 1994* (pp. 31-34). Hong Kong: University of Hong Kong.

Beemer,H. (1993). *Personal communication Bray, M., P. B. Clarke, & D. Stephens*. (1986). Education and Society in Africa. London: Edward Arnold.

Bentler, P. M. & Chou, C. (1987). Practical issues in structural modeling. *Sociological Methods and Research, 16*, 78-117.

Bentler P.M. (1992): Structural equation models in medical research. *Statistical Methods in Medical Research*, *1*(2), 159-181.

Biggs, J. B. (1995). Assumptions underlying new approaches to educational assessment. *Curriculum Forum, 4,* 1-22.

Blewchamp, P. (1994). *Washback in TOEFL classroom: an exploratory investigation into the influence of the TOEFL test on teaching content and methodology*. Unpublished MA thesis, University of Lancaster.

Bogdan, R. & Biklen, S. K. (1998). *Qualitative Research for Education: An Introduction to Theory and Methods.* London: Allyn & Bacon.

Borg, S. (2003). Teacher cognition in language teaching: a review of research on what language teachers think, know, believe, and do. *Language Teaching, 36*(2), 81–109.

Borg, S. (2006). *Teacher cognition and language education: Research and practice.* London: Continuum.

Bracey, G. W. (1987). Measurement-driven instruction: Catchy phrase, dangerous practice. *Phi Delta Kappan*, 683-686.

Breen, M.P. & Candlin, C. (1980). The essentials of a communicative curriculum in language teaching. *Applied Linguistics*, *1*(2), 89-112.

Brewer, J. & Hunter, A. (1989). *Multimethod research: a synthesis of styles.* Newbury Park, California: Sage.

Boyle, J., & Falvey, P. (Eds.). (1994). English language testing in Hong Kong. Hong Kong: Chinese University Press.

Brindley, G. (1989). *Assessing achievement in the learner-centred curriculum.* Sydney: Macquarie University, National Centre for English Language Teaching and Research.

Brindley, G. (2001). Assessment. In Carter, R. & Nunan, D. (Eds.), *Teaching English to speakers of other languages.* pp 137-143. Cambridge: CUP.

Boote, D. N., & Beile, P. (2005). Scholars before researchers: On the centrality of the dissertation literature review in research preparation. *Educational Researcher, 34*(6), 3–15.

Brooke, N., & J. Oxenham. (1984). The Influence of Certification and Selection on Teaching and Learning. In J. Oxenham (Ed.), *Education Versus Qualifications? A Study of Relationships Between Education, Selection for Employment and the Productivity of Labour.* London: Allen & Unwin.

Brookhart, S. & Freeman, D. (1992). Characteristics of Entering Teacher Candidates. *Review of Educational Research*, *62*(1), 37-60.

Broukal, M., & Nolan-Woods, E. (1995). *NTC's preparation for the TOEFL.* Lincolnwood, IL: National Textbook Company.

Brown, H. D. (1994). *Teaching by Principles: An Interactive Approach to Language Pedagogy.* N. J. : Prentice Hall Regents.

Brown, J. D. (1995). *The Elements of Language Curriculum*, USA: Heinle & Heinle.

Brown, J. D. (1996). *Testing in language programs.* Upper Saddle River, NJ: Prentice Hall Regents.

Brown, J. D. (1997). Designing surveys for language programs. In D. Griffee & D. Nunan (Eds.),*Classroom teachers and classroom research,* pp. 109-122. Tokyo: The Japan Association for Language Teaching (JALT).

Brown, H. D. (2003). *Language assessment: Principles and classroom practices.* White Plains, NY: Longman.

Brown, J. D. (2000). Extraneous variables and the washback effect. *Shiken: JALT Testing*, *3*, 21-33.

Brown, J. D. (2002). Questions and answers about language testing statistics: Extraneous variables and the washback effect. *Shiken: JALT Testing & Evaluation SIG Newsletter*, *6*(2), 12-.15

Brown, H. D. (2007). *Teaching by principles: An interactive approach to language pedagogy* (3rd ed.). New York: Pearson Longman.

Brown, J. D. & Hudson, T. (1998). *The Alternatives in Language Assessment. TESOL Quarterly*, *32*(4), 653-675.

Brumfit, C. (1991). Problems in defining instructional methodologies. In K. de Bot, R. P. Ginsberg, & Kramsch, C. (Eds.), *Foreign language research in cross-cultural perspective,* pp.133-144. Amsterdam/Philadelphia: John Benjamins.

Buck, G. (1988). Testing listening comprehension in Japanese university entrance examinations. *JALT Journal, 10,* 12-42.

Burnaby, B. & Sun, Y. (1989). Chinese Teachers' Views of Western Language Teaching: Context Informs Paradigms. *Tesol Quarterly, 23*(2), 219-238.

Burns, A. (1990). Focus on language in the communicative classroom. In G. Brindley (Ed.). *The Second Language Curriculum in Action.* Sydney: NCELTR.

Burrows, C. (1998) *Searching for washback: an investigation of the impact on teachers of the implementation into the Adult Migrant English Program of the assessment of the Certificates in Spoken and Written English.* Ph.D. thesis. Sydney: Macquarie Universit,Sydney.

Burrows, C. (2004). Washback in classroom-based assessment: A study of the washback effect in the Australian adult migrant English program. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 113-128). Mahwah, NJ: Lawrence Erlbaum Assoc.

Caine, A. Nicholas. (2005). *EFL Examination Washback in Japan: investigating the effect of oral assessment on teaching and learning.* MA thesis. England: The University of Manchester.

Cai, J. (2002). Pressures in current college English teaching. *Foreign Language Teaching and Research (Bimonthly)*, *34*(3), 228-230.

Cai, J. (2003). The Reasons for Carrying out Fundamental Reforms on Traditional College English Teaching. *China University Teaching*, *11*, 25-30.

Cai, J. (2005). Some Thoughts on College English Teaching. *Foreign Language*

Teaching and Research, *37*(2), 83-91.

Carless, D. (1999a). Factors affecting classroom implementation: Task-based curriculum renewal in Hong Kong. *International Journal of Educational Reform, 8*(4), 374-382.

Carless, D. (1999b). Large-scale curriculum change in Hong Kong. In Kennedy, C., Doyle, P., & Goh, C. (Eds.), Exploring Change in English Language Teaching. Oxford: Macmillan.

Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics, 1,* 1-47.

Candlin, C. and C. Edelhoff (1982). Challenges: Teachers' Book. Englewood Cliffs NJ: Prentice Hall.

Cantu, D. A. (2001). *An Investigation of the Relationship Between Social Studies Teachers' Beliefs and Practice*. Lewiston, New York: Mellen Press.

Campbell, K. P. & Zhao, Y. (1993). The Dilemma of English Language Instruction in the People's Republic of China. *Tesol Journal*, *2*(4), 4-6.

Cen, J. (1999). The status quo of foreign language teaching at the tertiary level. *Foreign Language Teaching and Research*, *117*(1), 3-6.

Chapelle, C., & Douglas, D. (1993). Foundations and directions for a new decade of language testing. In D. Douglas & C. Chapelle (Eds.), *A new decade of language testing research* (pp. 1-22). Alexandria, VA: TESOL.

Chapelle, C. A. (1998). Construct definition and validity inquiry in SLA research. In L. F. Bachman & A. D. Cohen (Eds.), *Interfaces between second language acquisition and language testing research*, pp. 32-70. New York: Cambridge University Press.

Chapman, D. W., & Synder, C. W. (2000). Can high-stakes national testing improve instruction: Reexamining conventional wisdom. *International Journal of Educational Development, 20,* 457-474.

Chen, G. (1999). How can the pot of English be brought to a boil? *Foreign Language World*, *74*(2), 26-29.

Chen, L. (2002). *Washback of a public examination: Impact of the Basic Competency Test on Taiwan junior high school English teaching*, presented in Edward F. Hayes Graduate. Research Forum and paper appeared in Edward F. Hayes Graduate Research. *Proceedings*, *2*, 91-99.

Chen, S. (1995). An Analysis of the CET 4 Backwash Effects. *Shantou University*

Journal (Humanities Quarterly), *2*(3), 78-82.

Chen, X. (1999). Reflection on Communicative Teaching in College English. *Journal of Chongqing University (Social Sciences Edition), 5*(3), 86-87.

Chen, Z. (2004). The need for respecting facts in academic discussions—Discussing with Mr. Runqing Liu on issues of the CET 4/6. *Foreign Language World*, *99* (1), 74-77.

Cheng, L. (1997). How does washback influence teaching? Implications for Hong Kong. *Language and Education, 11*(1), 3 8-54.

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. (1999). Changing assessment: washback on teacher perceptions and actions. *Teaching and Teacher Education,* 15, 253-271. *Curriculum Forum, 9(2)*, 1-33.

Cheng, L. (2000). Distance language teacher education: New challenges for Hong Kong. *Open Learning, 15*(1), 5-16.

Cheng, L. (2001). Washback studies: Methodological considerations. *Curriculum Forum*, *10*(2), 17-32.

Cheng, L. (2002). The washback effect on classroom teaching of changes in public examinations. In S. J. Savignon (Ed.), *Interpreting communicative language teaching: Contexts and concerns in teacher education* (pp. 91-111). New Haven, CT: Yale University Press.

Cheng, L. (2003). Looking at the impact of a public examination change on secondary classroom teaching: A Hong Kong case study. *Journal of Classroom Interaction, 38*(1), 1-10.

Cheng, L. (2004). The washback effect of a public examination change on teachers' perceptions toward their classroom teaching. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 146-170). Mahwah, NJ: Lawrence Erlbaum Associates.

Cheng, L. (2005). *Changing language teaching through language testing: A washback study. Studies in language testing, 21*. Cambridge: Cambridge University Press.

Cheng, L. (2008). The key to success: English language testing in China. *Language Testing, 25,* 15-37.

416

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

Cheng, L. (2008). Washback, impact and consequences. In E. Shohamy & N. H. Hornberger (Eds.), Encyclopedia of language and education, 2nd edition, Volume 7*: Language Testing and Assessment, 7, 349-364.*

Cheng, L. (2009). The history of examinations: Why, how, what and whom to select? InL. Cheng & A. Curtis (Eds.), *English language assessment and the Chinese learner* (pp. 13-25). New York and London: Taylor & Francis Group.

Cheng, L. (2009). *Teacher perspectives and actions toward a public examination change*. Unpublished manuscript. Hong Kong: University of Hong Kong, Department of Curriculum Studies.

Cheng, L., & Falvey, P. (2000) What works? The washback effect of a new public examination on teachers' perspectives and behaviors in classroom teaching. *Curriculum Forum, 9*(2), 10-33.

Cheng, L., & Curtis, A. (2004). Washback or backwash: A review of the impact of testing on teaching and learning. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (p. 3-17). Mahwah, NJ: Lawrence Erlbaum Associates.

Cheng, L. & Curtis, A. (2009). The realities of English language assessment and the Chinese learner in China and beyond. In L. Cheng & A. Curtis (Eds.), *English language assessment and the Chinese learner* (pp. 3-12). Routledge, New York and London: Taylor & Francis Group.

Cheng, L., Watanabe, Y., & Curtis, A. (2004). *Washback in language testing: Research contexts and methods*. Mahwah, NJ: Lawrence Erlbaum Associates.

Cheng, L. & Qi, L. (2006). Description and Examination of the National Matriculation English Test, *Language Assessment Quarterly*, *3*(1), 53-70.

Cheng, Sh. (1997). On College English Test Band IV and Band VI. *Journal of Guangzhou Normal University (Social Science Edition)*, *3*(3), 94-100.

Cheng, Sh. (2004). The dilemma caused by the norm. *Journal of Guangzhou University* (Social Science Edition), *3*(1), 88-96.

Chew, P.G.L. (2005, March). Change and continuity: English language teaching in Singapore. *Asian EFL Journal,* (7) 1. Retrieved from http://www.asian-efl-journal.com/ march_05_pc.php on December 24, 2010.

Cohen, A. D. (1984). On taking tests: What the students report. *Language Testing 1*(1), 70-81.

Cohen, A. D. (1994). *Assessing language ability in the classroom (2nd ed.).* New York: Heinle and Heinle.

Cohen, L., & Manion, L. (1994). *Research methods in education.* London: Routledge.

Choi, Inn-Chull (2008). The impact of EFL testing on EFL education in Korea. *Language Testing*, *25*(1), 39-62.

Chomsky, N. (1965). *Aspects of the theory of syntax.* Cambridge: CUP.

Chu, L., & Gao, P. (2006). An empirical study of the washback of CET-4 Writing. *Sino-US English Teaching, 3*(5), 36-38.

Cochran, W.G. (1977). *Sampling techniques. (3rd).* New York: Wiley & Sons., inc.

Cochran, W. G., & Cox, G. M. (1957). *Experimental Designs.* New York: John Wiley & Sons.

Cooksey, B. (1986). Policy and Practice in Tanzanian Secondary Education since 1967. *International Journal of Educational Development*, 6, 183-202.

Clark, J. L. D. (1983). Language testing: Past and current status--directions for the future. *Modern Language Journal, 67*(4), 431-443.

Clark, J. L. (1987). *Curriculum renewal in school foreign language learning.* Oxford, England: Oxford University Press.

Clift, R. (1987). English teacher or English major: Epistemological differences in the teaching of English. *English Education*, *19*, 229-236.

Colby-Kelly, C. & Turner, C.E. (2007). AFL research in the L2 classroom: taking assessment to the next level. *Canadian Modern Language Review*, *64* (1), 9-38.

Cotterall, S. (1995). Developing a course strategy for learner autonomy. *ELT Journal, 49*(3), 219-227.

Cook, T. D. & Campbell, D.T. (1979). *Quasi-experimentation: Design and analysis issues for field settings.* Boston: Houghton Mifflin.

Cooper, J. (1990). Cooperative Learning and College Teaching: Tips from the Trenches. *Teaching Professor*, *4*(5), 1-2.

Crandall, J. (2000). Language teacher education. *Annual Review of Applied Linguistics*, *20,* 34-55.

Creswell, J. W. (1994). *Research Design: Qualitative and Quantitative Approaches* (2nd ed.). Thousand Oaks, Calif: Sage Publications.

Creswell, JW. (2003). Research Design: Qualitative and Quantitative and Mixed Method Approaches. Thousand Oaks, CA: Sage Publications.

Creswell, J & Plano Clark, V.L. (2007). *Designing and conducting mixed methods*

418

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

research. Thousand Oaks , CA : SAGE.

Creswell, J. (2009). *Research design : qualitative, quantitative, and mixed methods approaches.* Thousand Oaks, CA: Sage.

Crook, D. (1998). *Some Problems of Chinese Education as Seen Through the Eyes of a Foreigner.* Beijing: Foreign Language Teaching and Research Press.

Cronbach, L. J. (1970). *Essentials of psychological testing* (3rd ed.). New York: Harper & Row.de Jong, J. H. A. L., & Stevenson, D. K. (Eds.). (1990). *Individualizing the assessment of language abilities.* Clevedon: Multilingual Matters.

Cunningsworth, A. (1995). *Choosing Your Course book.* Heinemann.

Dai, W. (2001). The construction of the streamline ELT system in China. *Foreign Language Teaching and Research*, *33*(5), 322-327.

Davison, C. (2008). *Using summative assessments for formative purposes: The ultimate justification for learners and teachers.* Presented at 30th Annual Language Testing Colloquium, (LTRC 08). Hangzhou.

Davison, C. & Hamp-Lyons (2008). *Applying school-based assessment for learning in a formal examination system.* Workshop presented at the 30th Language Testing Research Colloquium, Hangzhou, China.

Davies, A. (1990). *Principles of language testing*. Oxford: Basil Blackwell Ltd.

Davies, A. (1968). *Language testing symposium: A psycholinguistic approach*. Oxford: Oxford University Press.

Davies, A. (1985). Follow my leader: Is that what language tests do? In Y. P. Lee, A. C. Y. Y. Fok, R. Lord, & G. Low (Eds.), *New directions in language testing* (pp. 3-12). Oxford: Pergamon Press.

Davies, J.A. (1971). Elementary survey analysis. Englewood, NJ: Prentice-Hall

Davies, P. M. & Pearse, E. (2000) *Success in English Teaching*. Oxford: Oxford University Press.

Denscombe, M. (2007) *Communities of practice and mixed methods research: a research paradigm based on reflexive pragmatism.* Paper presented at International Mixed Methods conference, Cambridge, July 2007.

Denzin, N. K. (1970). *Sociological methods: A source book*. Chicago, IL: Aldine.

DeVincenzi, F. 1995 "Language Tests and ESL Teaching – Examining Standardised Tests Content: Some Advice for Teachers". *TESOL Quarterly, 29*(1) 180-184.

419

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

Dickinson, L. (1982). *Self-assessment as an aspect of autonomy*. Edinburgh: Scottish
    Centre for Overseas Education.

Dickinson, L. (1987). *Self-instruction in language learning*. Cambridge: Cambridge
    University Press.

Donnelly, W. & Yu, Zh. (2002). Testing the Test: CET and Communicative Validity.
    Journal of Chendu University of Information Technology, *17*(2), 138-148.

Dong, Y. (2003) Reading & writing-Top priority in China's college English education.
    Foreign Language World, *1*, 2-6.

Drenth, P. J. D., H. van der Flier, & I. M. Omari.(1983). Educational Selection in
    Tanzania. *Evaluation in Education*, 7, 95.

Dunkel, P. (Ed.). (1991). Computer-assisted language learning and testing: Research
    issues and practice. New York: Newbury House.

Eckstein, M. A. and Noah, H. J. (Eds.) (1993a). *Examinations: Comparative and
    international studies.* Pergamon Comparative & International Education Series.
    Oxford: Pergamon Press.

Eignor, D., Taylor, C., Kirsch, I., & Jamieson, J. (1998). Development of a scale for
    assessing the level of computer familiarity of TOEFL examinees TOEFL
    Research *Report No. 60.* Princeton, NJ: Educational Testing Service.

Eisemon, T. 0. (1990). Examination Policies to Strengthen Primary Schooling in
    African Countries.*International Journal of Educational Development*, *10*, 69

Elley, W. B. (1992). *How in the World do Students Read?* The Hague: International
    Association for the Evaluation of Educational Achievement.

Ellis, R. (1997). The Empirical Evaluation of Language Teaching Materials. *ELT
    Journal, 51* (1), 5-31.

Ellis, R. (2001). *The study of second language acquisition.* Oxford: Oxford University
    Press.

Elder, C. & K. O'Loughlin (2003) Score gains on IELTS after 10-12 weeks of intensive
    English study. *IELTS Research Report*, *4,* 62-87.

Elder C. (2001). Assessing the language proficiency of teachers: Are there any border
    controls? *Language Testing*, *18*(1), 149-170.

English, Fenwick W. (2000). *Deciding what to teach and test: developing, aligning,
    and auditing the curriculum.* Thousand Oaks, CA: Corwin Press, Inc.

Enright, M. K. (2004). Research issues in high-stakes communicative language testing:
    Reflections on TOEFL's new directions. *TESOL Quarterly, 38*(1), 147-151.

Erfan, N. (1990). *Examinations in Pakistan: An Overview.* Report of the National
Workshop on Techniques of Paper Setting and Marking. Lahore: Curriculum
Research and Development Center.

Ernest P. (1989). The knowledge, beliefs and attitudes of the mathematics teacher: A
model. *Journal of Education for Teaching*, *15*(19), 13-33.

Fang, Z. (1996). A review of research on teacher beliefs and practices. *Educational
Research*, *38,* 47-65.

Feng, Q. (2000) A discussion of the problems of College English teaching and reform
strategies. *Foreign Language Teaching, 3,* 55-61.

Ferman, I. (2004). The washback of an EFL national oral matriculation. In L. Cheng,
Y. Watanabe, & A. Curtis (Eds.), Washback in language *testing: Research
contexts and methods* (pp. 191-210). Mahwah, NJ: Lawrence Erlbaum
Associates.

Finnochiaro, M., & Sako, S. (1983). *Foreign language testing: A practical approach.*
New York: Regents Publishing Company.

Fournier-Kowaleski, L. (2005). *Depicting Washback in the Intermediate Spanish
Language Classroom: A Descriptive Study of Teacher's Instructional
Behaviors as They Relate to Tests.* Ph.D. Thesis*.* The Ohio State University

Fraenkel, J. R. & Wallen, N. E. (2003). *How to design. and evaluate research in
education.* Retrieved October 17, 2010 from http://static.toodoc.com/view

Freeman, D. (1996). Redefining the relationship between research and what teachers
know. In K. M. Bailey & D. Nunan (Eds.), *Voices from the language
classroom: Qualitative research in second language education*. Cambridge:
Cambridge University Press.

Freeman, Y.S. & Freeman, D.E. (1992). *Whole language for second language
learners*. Portsmouth, NH: Heinemann.

Freeman, D. (1996). Renaming experience/ reconstructing practice: Developing new
understandings of teaching. In Freeman, D. & Jack C. Richards (Eds.), *Teacher
learning in languageteaching*, pp. 221–41. New York:Cambridge University
Press.

Freeman, D. (2002). The hidden side of the work: teacher knowledge and learning to
teach. A perspective from North American educational research on teacher
education in English language teaching. *Language Teaching*, *35*, 1–13.

Freeman, D., Johnson, K.E. (1998). Reconceptualizing the knowledge-base of

language teacher education. *TESOL Quarterly, 32*(3), 397-417.

Freeman, D. & Richards, J. C. (1993). Conceptions of teaching and the education of second language teachers. *TESOL Quarterly, 27*(2), 193-216.

Freeman, D. & J. Richards (1996). *Teacher learning in language teaching.* Cambridge: Cambridge University Press.

Fredericksen, N. (1984). The real test bias: Influences on testing and teaching. *American Psychologist, 39*, 193-202.

Frederiksen, J. R., & A. Collins. (1989). A Systems Approach to Educational Test-ing. *Educational Researcher, 18,* 27-32.

Fullilove, J. (1992). The tail that wags. *Institute of Language in Education Journal, 9,* 131-147.

Fullan, M.D. (1982). *The meaning of educational change*. NY: Teachers College Press.

Fullan, M. (1993). *Change forces: Probing the depths of educational reform.* Bristol, PA: Falmer.

Fullan, M. (1999). *Change Forces: The Sequel.* (1st ed.). London: Falmer Press.

Fullan, M. (2001). *Leading in a culture of change*. San Francisco: John Wiley an Sons.

Fullan, M.G. & Stiegelbauer, S. (1991). *The New Meaning of Educational Change.* (2nd ed.). New York: Teachers College Press.

Gao, Y. (2003). Need for Scientific Evidence in Foreign Language Teaching and Learning Reforms. *Foreign Language Teaching and Research*, *3,* 222-223.

Garson, (2007). Factor Analysis. Retrieved 11 January 23, 2007, from: http://www2.chass.ncsu.edu/garson/pa765/factor.htm.

Gardner, D. (1996). Self-assessment for self-access learners. *TESOL Journal, 5*(3), 18-23.

Gass, S. M., & Selinker, L. (2001). *Second language acquisition: An introductory course.* (2nd ed.). Mahwah, NJ: Erlbaum.

Gatbonton, E. & Gu, G. (1994). Preparing and Implementing a Task-based ESL Curriculum in an EFL Setting: Implications for Theory and Practice. *Tesol Canada Journal*, *11*(2), 9-29.

Gates, S. (1995). Exploiting washback from standardized tests. In J. D. Brown and S. O. Yamashita (Eds.), *Language Testing in Japan* (pp. 101-112). Tokyo: Japan Association for language Teaching.

Gibbons, P. (2002). *Scaffolding language, scaffolding learning: Teaching second language learners in the mainstream classroom.* Portsmouth NH: Heinemann.

Gibbons, P. (2006). *Bridging discourses in the ESL classroom.* London: Continuum.

Gillham, B. (2005). *Researching interviewing: the range of techniques.* Maidenhead: Open University Press.

Gipps, C., & P. Murphy. (1994). A Fair Test? Assessments, Achievement and Equity.

Glaser, R. and Bassok, M. (1989). Learning Theory and the Study of Instruction. *Annual Review of Psychology, 40*, 631-666.

Glaser, R. and Silver, E. (1994). Assessment, Testing, and Instruction: Retrospect and Prospect. *Review of Research in Education, 20,* 393-419.

Glaser, B.G., and Strauss, A.L. (1990). *The Discovery of Grounded Theory: Strategies for Qualitative Research.* Chicago: Aldine.

Glesne, C., & Peshkin, A. (1992). *Becoming qualitative researcher: An introduction.* White Plains, NY: Longman

Gordon, W. (1988).Malawi Higher Education; Problems, Needs and Priorities. Washington DC: World Bank

Graves, K. (2000) *Designing Language Courses*. Canada: Heinle & Heinle..

Greaney, V. (1991). *Sudan: Public Examinations*. Unpublished paper. Washington DC: World Bank.

Green, D. Z. (1985). Developing measures of communicative proficiency: A test for French immersion students in grades 9 and 10. In P. Hauptman, R. LeBlanc, & M. Wesche (Eds.), *Second language performance testing* (pp. 215-227). Ottawa: University of Ottawa Press.

Green, A. (2006). Watching for Washback: Observing the Influence of the International English Language Testing System Academic Writing Test in the Classroom. *Language Assessment Quarterly, 3*(4)*,* 333-368.

Green, A. (2007). IELTS washback in context. Cambridge: Cambridge University Press.

Green, A. and C. J. Weir (2005). Monitoring score gain on the IELTS academic module in EAP programmes of varying duration. *IELTS Research Reports.*

Greene, R. and Wall, D. (2005). Language testing in the military: problems, politics and progress. *Language Testing, 22*(3), 379–398.

423

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

Green, A. (2003). Test impact and English for academic purposes: A comparative study in backwash between IELTS preparation and university presessional courses. Unpublished doctoral dissertation, University of Surrey, Roehampton, UK.

Greene, J. C. (2007). *Mixed methods in social inquiry*. San Francisco, CA: Jossey.

Greene, J. C., Caracelli, V. J., & Graham W. F. (1989). Toward a Conceptual Framework for Mixed-method Evaluation Designs. *Educational Evaluation and Policy Analysis, 11* (3), 255-274.

Greene, JC, & Caracell, VJ. (Eds). (1997). Advances in Mixed Methods Evaluation: The Challenges and Benefits of Integrating Diverse Paradigms. San Francisco: Jossey-Bass.

Grossman, P. (1990). *The making of a teacher: Teacher knowledge and teacher education*. New York: Teachers College Press.

Grossman, P., & Richert, A. (1988). Unacknowledged knowledge growth: A re-examination of the effects of teacher education. *Teaching and Teacher Education, 4*(1), 53-62.

Gu, X. (2003). Counter-evidence of the Comment "Time-consuming with Low efficiency" on college English teaching and learning. *Foreign Language World*, *3,* 49-53.

Gu, X (2005) 'Positive or negative? An empirical study of CET washback on college English teaching and learning in China'. Unpublished doctoral dissertation, Shanghai Jiaotong University, China.

Guo, Y. (2003). Evaluation of College English Test. *Journal of Donghua University*, *20*(3), 144-146.

Gu, X. D. (2007). The empirical study of CET washback on college English teaching and learning in China. *Journal of Chong Qing University (Social Science Edition), 13*(4), 119-125.

Guo, S. (1999). *Current issues in teacher education in China.* Paper presented at the Annual Meeting of the Comparative and International Education Society. Toronto, Ontario, Canada.

Gunawardena, C. (1993). Employer Expectations and Equity in Education in Sri Lanka. *International Journal of Educational Development, 13*,125-130.

Gunter, M. A., Estes, T. H., & Schwab, J. (2003). *Instruction: A models approach* (4th ed.). Upper Saddle River, NJ: Pearson.

Guskey, T. R., & E. W. Kifer. (1990). Ranking m School Districts on the Basis of
State-Wide Test Results: Is it Meaningful or Misleading? *Educational
Measurement Issues and Practice, 9*, 11-16.

Haas, G. J. (1990). English language testing: The view from the admissions office. In
D. Douglas (Ed.), *English language testing in U.S. colleges and universities*.
Washington, DC: National Association for Foreign Student Affairs.

Hagiwara, P. (1983). Student placement in French: Results and implications. *Modern
Language Journal, 67*(1), 23-32.

Hale, G. A., Stansfield, C. W., & Duran, R. P. (1984). *Summaries of studies involving
the Test of English as a Foreign Language, 1963-1982* (TOEFL Research
Report 16). Princeton, NJ: Educational Testing Service.

Hammersley, M., & Atkinson, P. (1983). *Ethnography*: Principles and practice.
London: Routledge.

Hamp-Lyons, L. (1996). *Ethical test preparation: The case of TOEFL.* Paper presented
at the Language Testing Research Colloquium, Chicago, IL.

Hamp-Lyons, L. (1997). Washback, impact and validity: ethical concerns. *Language
Testing*, *14*(3), 295 – 303.

Haney,W. M. (1993). *Cheating and Escheating on Standardized Tests*. Paper
presented at Annual Meeting of the American Educational Research
Association, Atlanta GA, April 12-16.

Han, Baocheng. Dai, Manchun & Yang, Lifang. (2004). On the Problems of CET,
Foreign Languages and Their Teaching, *2*, 17-23

Han, C. (1999). An analysis of the practicality of the CLT. *Foreign Language World*,
74 (2), 8-10.

Han, B. (2002).The Urgent Need for Carrying out Testing Reforms on the Language
Proficiency of Chinese Tertiary-level Students. *Foreign Language Teaching
and Research*, *6*, 410-411.

Hao, Xiaojiang. (1993). Term Examinations, General Graduation Exarninations and
College Entrance Examinations. *Chinese Education and Society,* 26,37-46.

Harper, A. E., &V. S. Misra. (1976). *Research on Examinations in India*. New Delhi:
National Council of Educational Research and Training.

Hart, C. (1999). *Doing a literature review. Releasing the social science imagination*.
London: Sage Publications.

Hart, D., Lapkin, S., & Swain, M. (1987). Communicative language tests: Perks and perils. *Evaluation and Research in Education 1* (2), 83-93.

Hamp-Lyons, L. (1997). Washback, impact, and validity: Ethical concerns. *Language Testing, 14*(3), 295-303.

Hamp-Lyons, L. (1998). "Ethical Test Preparation Practice: The Case of the TOEFL". *TESOL Quarterly,* 33(2), 329-337.

Hamp Lyons, L. & Tavares, N. (2008). Retrieved on 6 May 2010 from http://sba.edu.hku.hk/new_qef/res_interactive.html.

Harvey, P.A. (1985).Lesson to be learned: Chinese approaches to language learning. *ELT Journal*, *39*(3), 183-186.

Hawes, H. (1979). *Curriculum and reality in African primary schools*. Harlow, England: Longman.

Hawkey, R. 2002. Impact studies as action research: The role of the teachers. 36th International Annual IATEFL Conference, University of York. (March 2002).

Hawkey, R. (2004). An IELTS impact study: Implementation and some early findings. *Research Notes, 15*. Cambridge: Cambridge ESOL.

Hawkey, R. (2006). Teacher and learner perceptions of language learning activity. *English Language Teaching Journal, 60*(3), pp. 242-252.

Hawkey, R.. (2009) *Consequential validity, in Taylor and Weir (eds) Examining Speaking*. Cambridge: Cambridge University Press and Cambridge ESOL.

Hawkey, R.A.H. (2009). *A study of the Cambridge Proficiency in English (CPE) exam washback on textbooks, in The Social and Educational Impact of Language Assessment*. Cambridge: Cambridge University Press and Cambridge ESOL.

Hayes, Belinda, M.(2003) *IELTS Preparation in New Zealand: an Investigation into the Nature of the Courses and Evidence of Washback.* Victoria University of Wellington: New Zealand. Retrieved on June 25, 2010 from http://researcharchive.vuw.ac.nz/

H. Torrance (1992). Changing Educational Assessment: International Perspectives and Trends. London: Routledge.

Herz, B., K. Subbarao, M. Habib & L. Raney. (1991).*Letting Girls Learn: Promising Approaches in Primary and Secondary Education*. Washington DC: World Bank.

426

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

Herman, J. L. and Golan, S. (1991). *Effects of standardized testing on teachers and learning-Another look*. CSE Technical Report 334. Los Angeles: University of California. Center for Research on Evaluation, Standards and Student Testing.

Herman, J. L., & Golan, S. (1993). The effects of standardized testing on teaching and schools. *Educational Measurement: Issues and Practice*, *12*, 20-42.

Herman, J. L. (1992). Accountability for testing. *R&D Exchange, 4*(3), 17-31.

Herman, J. L. and Golan, S. (1993). The effects of standardized testing on teaching and schools. *Educational Measurement: Issues and Practice*, *12*, 20-42.

Henning, G. (1987). A guide to language testing." Development, evaluation, research, lqew York: Newbury House.

Heyneman, S. P., & W. A. Loxley. (1983). The Effect of Primary-School Quality on Academic Achievement Across Twenty Nine High- and Low-Income Countries. *American Journal of Sociology, 88,* 1162-1194.

Heyneman, S. P. (1987). Uses of examinations in developing countries: Selection, research, and education sector management. *International Journal of Educational Development 7*(4*),* 251-263.

Heyneman, S. P., & Ranson, A, W. (1990). Using examinations and testing to improve educational quality. *Educational Policy, 4*, 177-192.

Higgenbotham, J. B. and Cox, K. K. (1979). *Focus group interviews*. Illinois: American Marketing Association.

Hilke, R., & Wadden, P. (1996). The TOEFL and its imitators: Analyzing the TOEFL and evaluating TOEFL-prep texts. *PASAA: A Journal of Language Teaching and Learning in Thailand*, *26,* 53-67.

Ho, Ping-ti. (1964). *The Ladder of Success in Imperial China. Aspects of Social Mobility 1368-1911*. New York: Columbia University, 1964.

Hoque, E. (2008). *English Language Teaching and Learning at the Alim Level in the Madrashas in Bangladesh: Problems and Possible Solutions*. M. Phil Thesis. Bangladesh. Jahangirnagar University.

Hsu, Hui-Fen (2009). *The impact of implementing English proficiency tests as a graduation requirement at Taiwanese universities of technology*. Ph.D. thesis. University of York. Retrieved January 27, 2010 from -etheses.whiterose.ac.uk

Huberman, M., & Miles, M. (1984). *Innovation up close*. New York: Plenum.

Huang, Ching-Ling . (2007). A Study of the Washback Effects of the Elementary

Level General English Proficiency Test on EFL Teaching and Learning Graduate Student. Retrieved July 20, 2010 from http://ethesys. yuntech.edu.tw/ETD-db/ETD-search

Huang, D. (2002). *A preliminary investigation of CET-4 washback.* Paper presented at the International Conference on Language Testing and Language Teaching. Shanghai: Shanghai Jiao Tong University, 224-233.

Huang, D. & Jin, G.. (2001). Washback Effect of Language Testing. *Journal of Southwest Jiaotong University* (Social Sciences), *2*(3), 107-110.

Huang, S (2004). Washback Effects of the Basic Competence English Test on EFL Teaching in Junior High School in Taiwan. *Journal of Curriculum Studies, 36*(1), 101-115.

Huang, Liyan (2005).The Washback Effect of a High-stakes Oral Test *Journal of Basic English Education, 8*(5), 42-69.

Hughes, A. (1988). Introducing a needs based test of English language proficiency into an English medium university in Turkey in A. Hughes (Ed.), Testing English for University Study. *ELT Documents, 127,*134-153, Modern English Publications.

Hughes, A. (1989). *Testing for language teachers*. Cambridge: Cambridge University Press.

Hughes, A. (2003). International comparisons of intermediate skills. *Learning & Skills Research Journal. 6*(4) pp. 5-9.

Hwang, Hee-Jeong (2003). *The impact of high-stakes exams on teachers and students: a washback study of the university entrance exam at the secondary school level in South Korea.* McGill Library and Collection. Retrieved on August 2010 from http://.digitool.Library.McGill CA:80

Hymes, D. (1972). On communicative competence. In Pride, J. & Holmes, J. (Eds.), *Sociolinguistics.* Harmondsworth: Penguin.

Ingulsrud, J. E. (1994). An entrance test to Japanese universities: Social and historical contexts. In C. Hill, & K. Parry (Eds.), *From testing to assessment: English as an international language.* London: Longman.

Jin, Y. (2000). The washback effects of College English Test-Spoken English Test on teaching. *Foreign Language World, 118*(2): 56–61.

Jin, Y. (2000). The backwash of the CET-SET on teaching. *Foreign Language World, 4,* 56-61.

Jin, Y. (2006). Improvement of test validity and test washback – The impact study of the College English Test Band 4 and 6. *Foreign Language World, 6(1*), 65-73.

Jin, Y. (2010).The place of language testing and assessment in the professional preparation of foreign language teachers in China. *Language Testing, 27*(4), 555-584.

Jing, Sh. (1999). Why can't the pot of English teaching be brought to a boil? China Youth Daily.

Johnson R K (Ed.) (1989). *The Second Language Curriculum.* Cambridge: CUP

Johnson R K (1989). *A decision-making framework for the coherent language curriculum.* in Johnson R K (Ed.)

Johnson, K. (1992). The instructional decisions of pre-service ESL teachers: New directions for teacher preparation programs. In J. Flowerdew, M. Brock, & S. Hsia (Eds.), *Perspectives on second language teacher education*. Hong Kong: City Polytechnic of Hong Kong.

Johnson, F., & Wong, C. L. K. L. (1981). The interdependence of teaching, testing and instructional materials. In J. A. S. Read., (Ed.), *Directions in language testing* (pp. 277-302). Singapore: Regional Language Centre.

Johnson, K. E. & Golombek, P. R. (2002). (Eds.) *Teachers' narrative inquiry as professional development*. NY: Cambridge University Press.

Johnson, R. B. & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Education Researcher*, *33*(7), 14-26.

Jou, Cheng-hsuan. (2008). Perceptions of the Test of English for International Communication and Its Washback Effects: A Case Study at a Private University in Taiwan. Retrieved April 11, 2010 from http://etds.lib.hcu.edu.tw/theabs/english_site/

Kalaja, P. (1995). Student beliefs (or metacognitive knowledge) about SLA reconsidered. *International Journal of Applied Linguistics. 5*(2), 191-204.

Kalaja, P. (2003). Research on students' beliefs about SLA within a discursive Approach. In Kalaja, P., & Barcelos, A. M. F. (Eds.), *Beliefs about SLA: New Research Approaches,* pp. 87-108. Dordrecht: Kluwer Academic Publishers.

Karavas-Doukas, E. (1996). Using attitude scales to investigate teachers' attitudes to the communicative approach. *English Language Teaching Journal*, *50*(3), 187-198.

429

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

Karabulut, Aliye. (2007). Micro level impacts of foreign language test (university entrance examination) in Turkey: A washback study. M.A. Thesis. Iowa State University**,**

Kariyawasam,T. (1993). *Learning, Selection and Monitoring: Resolving the Roles of Assessment in Sri Lanka*. Paper presented at Conference on Learning, Assessment and Monitoring, International Centre for Research and Assessment, Institute of Education, University of London.

Katsumasa, S. (1997). Communicative language testing: Principles and problems. *English Review*, *12*, 3-24. Retrieved December 1, 2009, from http:// ci.nii.ac.jp/naid/110004 693443/en/

Kellaghan T, G. F Madaus & P W Airasian (1982). *The Effects of Standardized Testing*. London: Kluwen, Nijholf Publishing

Kellaghan, T. (1 990). Examination Systems in Africa. In P. Broadfoot, R. Murphy & Torrance (Eds.), *Changing Educational Assessment. International Perspectives and trends.* London: Routledge.

Kellaghan, T.(1992). "Examination Systems in Africa: Between Internalization and Indigenization." In M. A. Eckstein& H. J. Noah (Eds.*), Examinations: Comparative and International Studies.* Oxford: Pergamon.

Kellaghan, T. (1993). *The Impact of External Examinations on Student and Teacher Behavior*. Paper presented at Seminar on Policies Affecting Learning Outcomes Through Impacts on Teachers, Harvard University. Cambridge MA, June 28-July 2.

Kehaghan T., & Greaney, V. (1992). *Using examinations to improve education: A study of fourteen African countries*. Washington DC: The World Bank.

Kennedy, C. (1988). Evaluation of the management of change in ELT projects. *Applied Linguistics, 9(4),* 329-342.

Kennedy, C. (1987). Innovating for a change: Teacher development and innovation. *English Language Teaching Journal*, *41*(3), 163-170.

Kennedy, C. (1988). Evaluation of the management of change in ELT

Khaniya, T. R. (1990). The washback effect of a textbook-based test. *Edinburgh Working Papers in Applied Linguistics, 1,* 48-58.

Khaniya, T. R. (1990b). *Examinations as instruments for educational change: Investigating the washback effect of the Nepalese English exams*. Unpublished doctoral dissertation, University of Edinburgh.

Kim, J. O. & Mueller, C.W. (1978). *Introduction to factor analysis: What it is and how to do it* Newbury Park. CA: Sage University Press.

Kline, R. B. (2005). *Principles and practice of structural equation modeling.* (2nd ed.). New York: The Guilford Press.

Kong, Y. & Nie, J (2002). The compound dictation in the CET and its washback effect of on teaching and learning. *Foreign Language World*, *88* (2), 51-57.

Krejcie, R. V., & Morgan, D. W. (1970). Determining sample size for research activities. *Educational and Psychological Measurement*, *30*(1), 607-610.

Kruger, R.A. (1994). *Focus groups.* (2nd ed.). California: Sage Publications, Inc.

Lam, H. P. (1994). Methodology washback—an insider's view. In D. Nunan, R. Berry, & V. Berry (Eds.), Bringing about change in language education: Proceedings of the *International Language in Education Conference 1994* (pp. 83-102). Hong Kong: University of Hong Kong.

Larsen, J. W. (1987). *Computerized adaptive testing.* In K. M. Bailey, T. L. Dale, & R. T. Clifford (Eds.), *Language testing research: Selected papers from the 1986 Colloquium* (pp. 1-10). Monterey: Defense Language Institute.

Larsen-Freeman, Diane. 1986. T*echniques and principles in language teaching*. Oxford: Oxford University Press.

Latimer, G. David. (2009).*Washback effects of the Cambridge preliminary English test at an Argentinean bilingual school.* Argentina. Retrieved March 21, 2010, from http://www.edu.tw/bicer/ca2/313/ca231324.htm.

Laurier, M. (1991). What we can do with computerized adaptive testing...and what we cannot do! In S. Anivan (Ed.), *Current developments in language testing* (pp. 244-255). Anthology Series 25. Singapore: Regional Language Centre.

LeBlanc, R., & Painchaud, G. (1985). Self-assessment as a second language placement instrument. *TESOL Quarterly, 19(4),* 673-687.

Leedy, P. (1997). *Practical Research: Planning and Design (6th ed.).* New Jersey: Prentice-Hall, Inc.

Levene, H. (1960). *Robust Tests for Equality of Variances*, in Contributions to Probability and Statistics, ed. I. Olkin, Palo Alto, CA: Stanford Univ. Press.

Levy & Ellis. (2006). *Towards a Framework of Literature Review Process in Support of Information Systems Research*. Proceedings of the 2006 Informing Science and IT Education Joint Conference. Retrieved on February 2011, from http://www.scis.nova.edu

431

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

Lewis, J. (1990). Self-assessment in the classroom: A case study. In G. Brindley (Ed.), *The second language curriculum in action* (pp. 187-213). Sydney: Macquarie University, National Centre for English Language Teaching and Research.

Li, H. (2009). Are teachers teaching to the test?: A case study of the College English Test (CET) in China. *International Journal of Pedagogies & Learning*, *5*(1), 25 – 36

Li, X. (1990). How powerful can a language test be? The MET in China. *Journal of Multilingual and Multicultural Development, 11*(5), 393-404.

Li, Jiongying. (2002). Reflection on the problems of CET. *Foreign Language Teaching, 5*, 33-38.

Likert, R. (1932). A Technique for the Measurement of Attitudes. *Archives of Psychology*, *140*, 1-33.

Linda, A. Bond. (2003). Norm- and criterion-referenced testing. *Practical Assessment, Research & Evaluation*, *5*(2). Retrieved June 24, 2011 from http://PAREonline.net/getvn.asp

Lincoln, Y. S. & Guba, E. G. (1985). *Naturalistic Inquiry*. Beverly Hill, CA: Sage.

Lindvall, C.M., and Nitko, A.J. (1975*). Measuring pupil achievement and aptitude*. New York: Harcourt Brace Jovanovich.

Liu, R. & Dai, M. (2003). *The Status Quo of Chinese College English Teaching Reform and Research on Its Development Strategies* [M]. Beijing: Foreign Language Teaching and Research Press.

Liu, D. (1995). *English Language Teaching In Schools In China* . Amity Foundation Teachers Project –Teaching Resources in the Amity Teacher Project Handbook.

Liu, R. (2002). Issues on the Reform of English Curriculum – from a Discrete Curriculum to An Uniformed Curriculum requirement. *Foreign Language Teaching and Research, 6,* 403- 404.

Liu, X. Y. (1988). An analysis of factors that hinder TESOL innovations in China. *TESL Reporter*, *21,*70-77.

Liu, R. & Dai, M. (2003). *The Status Quo of Chinese College English Teaching Reform and Research on Its Development Strategies* [M]. Beijing: Foreign Language Teaching and Research Press.

Liu, M., Zhao, C. G., & Tang, F. L. (2007). The new TOEFL iBT and its preliminary washback effect. *Foreign Language Teaching Abroad, 117*, 55-62.

432

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

Locke, L.F., Spirduso, W. W., & Silverman, S. J. (2000). <u>Proposals that work</u>. Newbury Park, CA: Sage Publications, Inc

Long, M.H. & Sato, C.J. (1983) Classroom foreigner talk discourse: forms and functions of teachers' questions. In H.W. Seliger & M.H. Long (Eds.). *Classroom Oriented Research in Second Languages*. Rowley, MA: Newbury House.

Long, M. H. (1985). *Input and Second Language Acquisition Theory.* In S. M. Gass & C. G. Madden (eds.), Input in Second Language Acquisition, pp. 377-393. Rowley, MA: Newbury House.

Long, M. H. (1996). The Role of Linguistic Environment In Second Language Acquisition. In W.C. Ritchie & T. K. Bhatia (eds.), *Handbook of Second Language Acquisition*, pp. 413-468. San Diego: Academic Press.

Lopez, Alexis. 2005. Examining the Potential Washback of a new large-scale English proficiency test: Aligning tests, teaching, and learning. University Of Illinois at Urbana-Champaign. Ph.D. Thesis. Retrieved June 25, 2010 from http://proquest.umi.com/pqdlink?

Luo, X. & Xiao, Y. (2002). On the Reform of the CET. *Foreign Language Teaching and Research*, *34*(4), 294-299.

Lougheed, L. (1992). *TOEFL practice book* (2nd ed.). Englewood Cliffs, NJ: Prentice Hall Regents.

Lowe, P. (1988). Introduction. In P. Lowe & C. W. Stansfield (Eds.), *Second language proficiency assessment: Current issues* (pp. 1-10). Englewood Cliffs, NJ: Prentice Hall Regents.

Locke, L., Spirduso, W. W., & Silverman, S. J. (1999). *Proposals that work* (4th ed.). Thousand Oaks, CA: Sage Publications.

Lumley, T., & Stoneman, B. (2000). Conflicting perspectives on the role of test preparation in relation to learning? *Hong Kong Journal of Applied Linguistics, 5*(1), 50-80.

Luxia, Q. (2005). Stakeholders' conflicting aims undermine the washback functioning of a high-stakes test. *Language Testing, 22*, 142-173.

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

Madaus, G. (1990). *Testing as social technology*. Paper presented at the inaugural

annual Boisi Lecture in Education and Public Policy. Boston College, 6

December.

Madaus, G. (1988). Testing and the curriculum. In L.N. Tanner, *Cultural issues in

curriculum.* 87th Yearbook, Part I, National Society for the Study of Education.

Chicago: University of Chicago Press, p. 83-121.

Madaus & T. Kellaghan. (1991). Student Examination Systems in the European

Community: Lessons for the United States. In G. Kulm & S. M. Malcom (Eds.),

Science Assessment in the Service of Reform. Washington DC: American

Association for the Advancement of Science.

Maniruzzaman, M. and Hoque, E (2010). How does Washback Work on the EFL

Syllabus and Curriculum? A Case Study at the HSC Level in Bangladesh.

Language in India, 10(12), 1-40.

Maniruzzaman, M. (2010). Relations of Washback to the Teachers of English as a

Foreign Language at the SSC Level. *Bangladesh Journal of Public Administration*

XIX, II.

Maniruzzaman, M. and Sultana, S. (2011). Washback of Second/Foreign Language

Tests: an Appraisal. *Jagannath University Journal of Arts* , *1*.

Manjarres, N. B. (2005). Washback of the foreign language test of the state

examinations in Colombia: A case study. *Arizona Working Papers in SLAT, 12,*

1-19.

Markee, N. (1997). *Managing Curriculum Innovation*. Cambridge: Cambridge

University Press.

Markee, N. (1993). The diffusion of innovation in language teaching. *Annual Review of

Applied Linguistics, 13*, 229-243.

Marshall, C. and Rossman, G. (1989). Designing Qualitative Research. Newbury

Park, CA: Sage Productions.

Maxwell, J.A. (1996). *Qualitative Research Design - An Integrative Approach.* Thousand

Oaks, London: Sage.

McLaughlin B. (1987). *Theories of second-language learning*. London: Arnold.

McLaughlin B. (1990a). "Conscious vs. unconscious learning". *TESOL Quarterly*,

24, 617-634.

McLaughlin B. (1990b). Restructuring. *Applied Linguistics*, *11*, 1-16.

McNamara, T. (2000) *Language Testing*, Oxford University Press ,Oxford,. 24-33.

Mehrens, W. A., & Kaminsky, J. (1989). Methods for improving standardized test

434

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

scores: Fruitful, fruitless or fraudulent? *Educational Measurement: Issues and Practices, 8*, 14-22.

Merriam, S. B. (1998). *Qualitative research and case study applications in education.* San Francisco, CA: Jossey-Bass.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher, 23*, 13-23.

Messick, S. (1996). Validity and washback in language testing. *Language Testing, 13*(3), 241-256.

Miller, L.E. (1999). *Lecture note of Ag. Ed. 885 Research methods*. Columbus, Ohio: The Ohio State University.

Miles, M., & Huberman, A. (1994). *Qualitative data analysis: A sourcebook of new methods* (2nd ed.). Beverly Hills, CA: Sage.

Mitchell, R. (1988). *Communicative Language Teaching in Practice*. London: CILT.

Miyazaki, I. (1976). *China's Examination Hall*. The Civil Service Examinations of Impernal China. New York: Weather hill.

Mizutani, S. (2009). *The mechanism of washback on teaching and learning*. Ph.D. Thesis. New Zealand: The University of Auckland

Mohammadi, M. (2007). *High-Stakes Testing Washback: A Survey on the Effect of Iranian MA Entrance Examination on Teaching*. Presented at the conference on English Language Teaching and Literature (ELTL) supported by TELLSI, 2007. Retrieved January 27, 2010 from http://riau.academia.edu/Mojtaba

Morris, P. (1990). *Curriculum development in Hong Kong*. Hong Kong: Faculty of Education, The University of Hong Kong.

Morris L. (1996).*Interaction in the language classroom: Awareness, Autonomy and Authenticity*. Harlow: Longman,

Morrow, K. (1986). The evaluation of tests of communicative performance. In M. Portal (Ed.), *Innovations in language testing* (pp. 1-13). NFER/Nelson.

Morrow, K. (1991). Evaluating communicative tests. In S. Anivan (Ed.), *Current developments in language testing* (pp. 111-118). Anthology Series 25. Singapore: Regional Language Centre.

Morse, J. M., & Richards, L. (2002). *Readme first for a user's guide to qualitative methods.* Thousand Oak: Sage.

Mousavi , M. and Mehrdad Amiri (2009).*The Washback Effect of TEFL University*

Entrance Exam on Academic Behavior of Students and Professors. Retrieved September 23, 2010 from http://eltweekly.com.

Muñoz, A. & Álvarez, M. (2010). Washback of an oral assessment system in the EFL classroom. *Language Testing*, *27* (1), 33-49.

Nambiar, M. and S. Ransirini. (2006) *Exploring Washback: A Study of the Test Impact of MUET.* A paper presented at the MICELT conference held in Penang, Malaysia 0305 May, 2006.

NCTB (2001). *English for Today* for classes 11-12. Dhaka: NCTB.

NCTB (2007). *English Grammar and Composition.* Dhaka: NCTB.

Niedermeyer, F., and Yelon, S. (1981). Los Angeles aligns instruction with essential skills. *Educational Leadership*, *38*(8), 618-620.

Nespor, J. (1987). The role of beliefs in the practice of teaching. *Journal of Curriculum Studies*, *19*(4), 317-328.

Newman, I. (2000). A conceptualization of mixed methods: A need for an inductive/deductive approach to conducting research. *Journal of Research in Education*, *10*(1), 81-84.

Noah, H. J., & M. A. Eckstein. (1992). Comparing National Systems of Secondary School Leaving Examinations. In M. A. Eckstein & H. J. Noah (Eds.), *Examinations: Comparative and International Studies*. Oxford: Pergamon.

Noble, A. J. & Smith, M. L. (1994a). *Old and new beliefs about measurement-driven reform: "The more things change, the more they stay the same"* (CSE Tech. Rep. No. 373). Los Angeles: UCLA, National Center for Research on Evaluation, Standards, and Student Testing.

Noble, A. J., & Smith, M. L. (1994b). *Measurement-driven reform: Research on policy, practice, repercussion* (CSE Tech. Rep. No. 381). Los Angeles: UCLA, National Center for Research on Evaluation, Standards, and Student Testing.

Nga, NP (2005). Teacher Evaluation and its Development" in Nga, N.P. edition. Higher Education: *Quality and Evaluation*, p.17-47.

Nunan, D. (1988). *The learner-centred curriculum: A study in second language teaching.* Cambridge: Cambridge University Press.

Nunan, D. (1996a). Towards autonomous learning: Some theoretical, empirical and practical issues. In R. Pemberton, E. S. S. Li, W. Or, & H. Pierson (Eds.), *Taking control: Autonomy in language learning* (pp. 13-26). Hong Kong: Hong Kong University Press.

436

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

Nunan, D. (1996b). *Hidden voices: Insiders' perspectives on classroom interaction*. In
K. M. Bailey & D.

Nunan, D. (1997). Language teaching and research. In D. Griffee & D. Nunan (Eds.),
Classroom teachers and classroom research, pp. 13-22. Tokyo: The Japan
Association for Language Teaching (JALT).

Nunan, D. (1999). Second language teaching and learning. Boston: Heinle & Heinle.

Nunan, D. (2001). Tradition and change in the ELT curriculum. In W. Hu (Eds),
Foreign Language Teaching and Research, pp. 50-58. Beijing: Foreign
Language Teaching and Research Press.

Nunan (Eds.), *Voices from the language classroom: Qualitative research in second
language education* (pp. 41-56). Cambridge: Cambridge University Press.

Oller, J. W., Jr. (1979*). Language tests at school: A pragmatic approach.* London:
Longman.

Nyawaranda V. (1999) Doing a qualitative research study in Zimbabwe. *Bulletin of
Teacher Education*, *2*(1), 7-13

Oskaarson, M. (1980). *Approaches to self-assessment in foreign language learning*.
Oxford: Pergamon Press.

Pearson, I. (1988). Tests as levers for change. In D. Chamberlain and R. Baumgardner
(Eds.), *ESP in the classroom: practice and evaluation* (pp. 98-107). ELT
Documents 128. Modern English Publications.

Ozaki, S. (2009) *The impact of entrance examinations on education for international
understanding*. Nagoya, Japan: Sankeisha. Proficiency testing, student
achievement, and local educational practices. (2000). Legislative Office of
Education Oversight [On-line].

Pajares, M. F. (1992). Teachers' beliefs and educational research: Cleaning up a
messy construct. *Review of Educational Research*, *62*(3), 307-332.

Pan, Y. (2008, Spring). A critical review of five language washback studies from
1995–2007: Methodological considerations. *Shiken: JALT Testing and
Evaluation SIG Newsletter*, *12*(2), 2–14. Retrieved July 8, 2010, from
http://jalt.org/test/pan_1.htm

Pan, Y. (2009). *The social impact of English certification exit requirements in
Taiwan*. Paper presented at Language Testing and Research Colloquium 2009,
Denver.

Patton, M.Q. (1990). *Qualitative Evaluation and Research Methods* (2nd ed.).

NewburyPark, CA: Sage.

Patton, M.Q. (2002). *Qualitative research & evaluation methods* (3rd ed.). Thousand
Oaks, CA: Sage.

Pearson, I. (1988). *Tests as levers of change*. In D. Chamberlain & R.J.Baumgartner
(Eds.), ESP in the classroom: Practice and evaluation (pp. 98-107). ELT
Documents #128. London: Modem English Publications in association with the
British Council.

Pierce, B. N. (1992). Demystifying the TOEFL Reading Test. *TESOL Quarterly,
26*(4), 665-689.

Popham, W. J. (1987). The merits of measurement-driven instruction. *Phi Delta
Kappa, 68*, 679– 682.

Popham, W. J. (1991). Appropriateness of teachers' test preparation practices.
Educational Measurement: lssues and Practices, 10(1), 12-15.

Porter, A. (2002). Measuring the content of instruction: Uses in research and practice.
Educational Researcher, 31. 3-14.

Porter, A. (2004). Curriculum assessment. *Complementary Methods for Research in
Education*. J. Green, G. Camilli, & P. Elmore (Eds.) Washington DC: AERA.
pp. 141-159.

Prodromou, L. (1995). The backwash effect: From testing to teaching. *ELT Journal,
49*(1), 13-25.

Qi, L. (2004). Has a high-stakes test produced the intended changes? In L. Cheng, Y.
Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts
and methods* (pp. 171-190). Mahwah, NJ: Lawrence Erlbaum Associates.

Qi, L. (2005). Stakeholders' conflicting aims undermine the washback function of a
high-stake 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.

Qian, J. (2003). On the Impeding Factors of CET Reform and Development. *Journal of
Zhongzhou University*, *20*(4), 91-92.

Rabbe, L., & Shushter-Bouskila, G. (2001). *Embracing the different learner styles in
the EFL curriculum*. Retrieved July 3, 2009, from http://www.etni.org.il/etai

Re-Dickens, P.(1997) 'So, why do we need relationships with stakeholders in language
testing? A view from the UK' *Language Testing, 14*(3), 304-314

Rea-Dickens, P. (2004). Editorial: Understanding teachers as agents of assessment. *Language Testing*, *21*(3), 249 -258.

Read, J. and Hayes, B.(2003) The impact of IELTS on preparation for academic study in New Zealand. *IELTS International English Language Testing System Research Reports*, *4,* 153-206.

Read, J. and Hayes, B., (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: Research contexts and methods* (pp. 97-111). Mahwah, NJ: Lawrence Erlbaum Associates.

Raimes, A. (1990). The TOEFL Test of Written English: Causes for concern. *TESOL Quarterly, 24*(3), 427-442.

Rao, Z. (2001). Chinese students' perceptions of communicative and non-communicative activities in EFL classroom. *System*, *30,* 85-105.

Reagon, T. G. & Osborn, T. A. (2002). *The Foreign Language Educator in Society. Toward a Critical Pedagogy*. (1st ed.). Mahwah, New Jersey: Lawrence Erlbaum Sociates, Publishers.

Reardon, S. F., Scott, K. and Verre, J. (1994) Symposium: equity in educational assessment. *Harvard Educational Review*, *64* (1), 1–4.

Retorta, S. Miriam (2007). *The washback effect of the Federal University entrance* examination of Parana on the teaching of the English language in secondary schools of Parana. Retrieved April 25, 2009, from http://linguistlist.org

Richards, J. C. and Nunan, D. (1990). Second language teacher education. Cambridge: Cambridge University Press.

Richards, J. (2001) *Curriculum Development in Language Teaching,* Cambridge: Cambridge University Press.

Richards, J. (2008). Second Language Teacher Education Today. *RELC Journal*, *39*(2), 158-177.

Richards, J. (2006). Materials development and research: Making the connection. *RELC Journal, 37*(1), 5-26. Retrieved June 5, 2010, from EBSCOhost database.

Richards, J. (2001) *Curriculum Development in Language Teaching,* Cambridge: Cambridge University Press.

Richards, J. & Lockhart, C. (1994). *Reflective teaching in second language classrooms.* New York: Cambridge University Press.

439

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

Richards, J. C., & Farrell, T. (2005). *Professional development for language teachers: strategies for teacher learning*. Cambridge: Cambridge University Press.

Richards, J., & Rodgers, T. (2001). *Approaches and methods in language Teaching* (2nd ed.). Cambridge: Cambridge University Press.

Richardson, V. (1996). The role of attitudes and beliefs in learning to teach. In J. Sikula, T. J. Buttery, & E. Guyton (Eds.), *Handbook of research on teacher education* (2nd ed.), pp. 102-119. New York: Macmillan.

Robert, G. (1997). *Statistical Methods for Education*. New York: Prentice Hall.

Rogers, B. (1993). *The complete guide to the TOEFL*. Boston: Heinle and Heinle.

Rolfe, T. (1990). Self- and peer-assessment in the ESL curriculum. In G. Brindley (Ed.), *The second language curriculum in action* (pp. 163-186). Sydney: Macquarie University, National Centre for English Language Teaching and Research.

Ross, S. J. (2005). A book review of washback in language testing: research contexts and methods. In Cheng. L., Watanabe, and Curtis (eds.). *Applied Linguistics*, *26*(3), 459-463.

Rossman, GB. & Wilson, BL. (1985). Numbers and words: Combining qualitative and quantitative methods in a single large scale evaluation. *Evaluation Review,* *9*(5), 627-643.

Rossman, GB. & Wilson, BL. (1994). Numbers and words revisited: Being shamelessly eclectic. *Quality and Quantity*, *28*(3), 315-327.

Saif, S. (2006). Aiming for positive washback: A case study of international teaching assistants. *Language Testing*, *23*(1), 1-34.

Saif, S. (1999). *Theoretical and empirical considerations in investigating washback: A study of ESL/EFL learners*. Doctoral Dissertation, University of Victoria.

Samuelowicz, K., and Bain, J.D. (1992). Conceptions of teaching held by academic teachers, *Higher Education*, *24*, 93–111.

Samuelowicz, K., and Bain, J.D. (2001). Revisiting Academics' Beliefs about Teaching and Learning, *Higher Education*, *41,* 318.

Satterthwaite, F. W. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin, 2*, 110-114.

Sato K and Kleinsasser RC (1999) Communicative Language Teaching (CLT): Practical Understandings. *The Modern Language Journal, 83*(4),494-517.

Saville, N. (2000). Investigating the impact of international language examinations. *UCLES Research Notes* 2. [Online] Available: http://www.cambridge-efl.org.uk/rs_notes/0002/rs_notes2_3.cfm.

Saville, N., & Hawkey, R. (2004). The IELTS impact study: Investigating washback on teaching materials. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 73-96). Mahwah, NJ: Lawrence Erlbaum Associates.

Saville, N. (2008). *Assessment as a washback tool: is it beneficial or harmful?* University of Cambridge. ESOL Examinations. Retrieved on July 2010 from www.cambridgeesol.org/assets/pdf/

Scheaffer, Richard L., Mendenhall, William and Ott, Lyman. (1996). *Elementary Survey Sampling*. Wadsworth Publishing, Belmont, California.

Schutz, N. W., Jr., Derwing, B., Palmer, I., & Steed, J. F. (1991). *Breaking the TOEFL barrier*. Englewood Cliffs, NJ: Prentice Hall.

Sharpe, P. J. (1996). *Barron's students' #1 choice*. How to prepare for the TOEFL (eighth edition). Woodbury, NY: Barron's.

Shavelson, R J. and R Stern. (1981). Research on teachers' pedagogical thoughts, judgments, decisions, and behavior. *Review of Educational Research, 51*, 455-98.

Shaw, P. (1980). Comments on the concept and implementation of self-placement. *TESOL Quarterly, 14*(2), 261-262.

Sheldon, L. (1988). Evaluating ELT Textbooks and Materials. *ELT Journal, 42(2),*3-26.

Shepard, L.A. (1991). Psychometricians' beliefs about learning. *Educational Researcher*, *20*(6), 2–16.

Shepard, L.A. (1993). The place of testing reform in educational reform: A reply to Cizek. *Educational Researcher*, *22* (4), 10–14.

Shih, C. (2007). A new washback model of students' learning. Canadian Modern *Language Review*, *64,* 135-162.

Shih, C. (2008). Critical Language Testing: A case of the General English Proficiency Test. *English Teaching and Learning, 32*(3), 1-34.

Shih, C. (2009). How tests change teaching: A model for reference. *English Teaching: Practice and Critique, 8*(2), 188-206.

Shih, C. (2010). The washback of the General English Proficiency Test on university policies: A Taiwan case study. *Language Assessment Quarterly, 7*, 234-254.

Shohamy, E. (1992). Beyond proficiency testing: A diagnostic feedback testing model for assessing foreign language learning. *Modern Language Journal, 76*(4), 513-521.

Shohamy, E. (1992). New models of assessment: The connection between testing and learning. In E. Shohamy & R. Walton (Eds.), *Language assessment for feedback: Testing and other strategies.* Dubuque, IA: Kendall Hunt Publishing Company.

Shohamy, E. (1993). *The Power of Test: The Impact of Language Testing on Teaching and Learning.* NFLC Occasional Papers, ED 362-400.

Shohamy, E. (1993). A collaborative/diagnostic feedback model for testing foreign languages. In D. Douglas, & C. Chapelle (Eds.), *A new decade of language testing research* (pp. 185-202) . Alexandria, VA: TESOL Publications.

Shohamy, E. (1997). Testing Methods, Testing Consequences: Are they Ethical? Are they Fair? *Language Testing*, *14*(3), 340-49.

Shohamy, E. (2005). The power of tests over teachers: the power of teachers over tests. In D.J. Tedick (Ed.), *Second language teacher education: International perspectives* (pp. 101-111). Mahwah, NJ: Lawrence Erlbaum Associates.

Shohamy (2007). Tests as power tools: Looking forward, looking back. In J. Fox et al. (Eds.), *Language testing reconsidered* (pp. 141- 152). Ottawa: University of Ottawa Press.

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

Shohamy, E., Reves, T., & Bejarano, Y. (1986). Introducing a new comprehensive test of oral proficiency. *ELT Journal, 40*(3), 212-220.

Shulman, L.S. (1986a). Paradigms and research programmes in the study of teaching: a contemporary perspective, in M.C. Wittrock (Ed.) *Handbook of Research on Teaching,* pp. 505-526, (3rd ed.). New York: Macmillan.

Shulman, L. (1986b). Those who understand: knowledge growth in teaching, *Educational Researcher*, *15,* 4-14.

Shulman, L. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Education Review*, *57*(1), 1-21.

Shulman, L. S. (1992). On research on teaching: A conversation with Lee Shulman. *Educational Leadership*, 49(7), 14-19. Interviewer Ron Brandt.

Shulman, L. S. (2000). Teacher development: Roles of domain expertise and

pedagogical knowledge. *Journal of Applied Developmental Psychology*, *21*(1), 129-135.

Silva de Oliveira, J. A. (2009). *Washback effect of achievement testing in Brazilian regular education: keeping an eye on motivation to learn EFL*. Biblioteca Digital da UNICAMP. Retrieved October 2010, from http://cutter.unicamp.br/

Skehan, P. (1998). Task-based instruction. *Annual Review of Applied Linguistics*, *18,* 268-286.

Smallwood, I. M. (1994). Oral assessment: A case for continuous assessment at HKCEE level. *New Horizons: Journal of Education*, Hong Kong Teachers' Association, *35,* 68-73.

Smith, M. L. (1991). Put to the test: The effects of external testing on teachers. *Educational Researcher, 20*(5), 8-11.

Smith, M.L. (1991a). Meanings of test preparation. *American Educational Research*

Smith, M.L., Noble, A.J., Cabay, M., Heinecke, M., Junker, M.S. & Saffron, Y. (1994). *What happens when the test mandate change*? Results of a multiple case study. CRESST Report 380. Arizona, USA: Arizona State University.

Snyder C W Jr, B Prince, G Johanson, C Odaet, L Jaji & M Beatty (1997).*Exam Fervor and Fever: Case Studies of the Influence of Primary Leaving Examinations on Uganda Classrooms, Teachers and Pupils* (Volume One)' Advancing Basic Education and Literacy (ABEL) Project Academy for Educational Development, Washington DC .*Journal*, *28*(3), 521-542.

Solomon, P. (2002). *The assessment bridge: Positive ways to link tests to learning, standards, and curriculum improvement.* Thousand Oaks, CA: Corwin.

Spada, N. & Fröhlich, M. (1995). *The Communicative Orientation of Language Teaching Observation Scheme (COLT)*. The National Centre for English Language Teaching and Research (NCELTR). Macquarie University, Sydney, Australia.

Spolsky, B. (1975). Language testing – the problem of validation. In L. Palmer & B. Spolsky (Eds.). Papers on Language Testing 1967-1974. Washington, D.FC.: *TESOL*. 147-53.

Spolsky, B. (1985). What does it mean to know how to use a language? An essay on the theoretical basis of language testing. *Language Testing, 2* (2), p.180-191.

Spolsky, B. (1990). Social aspects of individual assessment. In J. H. A. L. de Jong & D. K. Stevenson (Eds.), *Individualizing the assessment of language abilities*. Clevedon: Multilingual Matters.

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.

Spratt, M. (2005). Washback and the classroom: The implications for teaching and learning of studies of washback from exams. *Language Teaching Research, 9*(1), 5- 29.

Stansfield, C. W. (Ed.). (1986). *Technology and language testing*. Washington, DC: TESOL Publications.

Stecher, B., Chun, T., & Barron, S. (2004). The effects of assessment-driven reform on the teaching of writing in Washington state. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 53-71). Mahwah, NJ: Lawrence Erlbaum Associates.

Steffy, B. E. (2001). *Deep curriculum alignment: Creating a level playing field for all children on high-stakes tests of educational accountability*. Lanaham, MD: The Scarecrow Press, Inc

Stephens, D., Pearson, P. D., Gilrane, C., Roe, M., Stallman, A., & Shelton, J. (1995). Assessment and decision making in schools: A cross-site analysis. *Reading Research Quarterly*, *30,* 478–499.

Stern, H.H. (1992). *Issues and options in language teaching*. Oxford: Oxford University Press.

Stobart, G. (2003). The impact of assessment: Intended and unintended consequences. *Assessment in Education, 16*, 139-140.

Stoller, F. (1994). The diffusion of innovations in intensive ESL programs. *Applied Linguistics 15*(3), 300-327.

Strauss, A., & Corbin, J. (1990): *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*. London: Sage.

Strauss, A. & Corbin, J. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (2nd ed.). Thousand Oaks: Sage.

Sturman, P. (1996). Registration and placement: Learner response. In K. M. Bailey & D. Nunan (Eds.), *Voices from the language classroom: Qualitative research in*

second language education (pp. 338- 355). Cambridge: Cambridge University Press.

Swain, M. (1984). Large-scale communicative testing: A case study. In S. L. Savignon & M. Berns (Eds.), *Initiatives in communicative language teaching* (pp. 185- 201). Reading, MA: Addison Wesley.

Swain, M. (1985). Large-scale communicative testing: A case study. In Y. P. Lee, A. C. Y. Fok, R. Lord, & G. Low (Eds.), *New directions in language testing* (pp. 35- 46). Oxford: Pergamon Press.

Stoneman, B. W. H. (2006). *The impact of an exit English test on Hong Kong undergraduates: A study investigating the effects of test status on students' test preparation behaviours.* Unpublished Ph. D. Dissertation, The Hong Kong Polytechnic University, Hong Kong, China.

Tan, M. (2008). *Bilingual high-stakes mathematics and science exams in Malaysia: pedagogical and linguistic issues.* Paper presented at the Language Testing and Research Colloquium 2008, Hangzhou.

Tang, X. Y. (2005). The washback study of language testing. *Foreign Languages and Their Teaching, 7*, 55-59.

Tashakkori, A. & Teddlie, C. (1998). *Mixed Methodology: Combining qualitative and quantitative approaches.* Thousand Oaks, CA: Sage.

Taylor L (2000). *Stakeholders in language testing*. University of Cambridge Local Examinations Syndicate (UCLES) Research Notes 2. Retrieved from http://www.cambridge-efl.org/rs_notes [last accessed 04.09.2010]

Taylor, C., Jamieson, J., Eignor, D., & Kirsch, I. (1998). *The relationship between computer familiarity and performance on computer-based TOEFL test tasks* (TOEFL Research Report No. 61). Princeton, NJ: Educational Testing Service.

Taylor, S. J. & Bogdan, R. (1998). Introduction to qualitative research methods: A guidebook and resource (3rd ed.). New York: John Wiley and Sons

Tavares, N. & Hamp-Lyons, L. (2008). *ESL teachers learning to assess interactively: The challenges and teacher development issues.* Presented at 30th Annual Language Testing Colloquium, (LTRC 08). Hangzhou.

Teddlie, C. & Tashakkori, A. (2009). *Foundations of mixed methods research: Integrating quantitative and qualitative techniques in the social and behavioral sciences.* Thousand Oaks, CA: Sage.

Thompson, A. (1992). Teachers' beliefs and conceptions: A synthesis of research. In

D.A Grouws (Ed.), *Handbook of research on mathematics teaching and learning*, pp. 127-146. New York: Macmillan.

Ting, Y. R. (1987). Foreign Language Teaching in China: Problems and Perspectives. *Canadian an International Education*, *16,* 48-61.

Tong, Keith, & Others (1991). *To Test or Not To Test, That Is the Question*. (ED354750) . Retrieved March 21, 2010 from www.cambridgeesol.org/assets/.

Tsagari, D. (2007*).* Investigating the Washback effect of the FCE examination (Cambridge/ESOL) in Greece. Greece ,UCLES.

Tsagari, D. (2009). *The Complexity of Test Washback: An Empirical Study*. Language Testing and Evaluation Series, Grotjahn, R. &. G. Sigott (eds). Frankfurt am Main: Peter Lang GmbH.

Tung, P. (1985). Computerized adaptive testing: Implications for language test developers. In P. C. Hauptman, R. LeBlanc, & M. B. Wesche (Eds.), *Second language performance testing* (p. 12-27). Ottawa: University of Ottawa Press.

Turner, C.E. (2001). The need for impact studies of L 2 performance testing and rating: Identifying areas of potential consequences at all levels of the testing cycle. In A. Brown et al. (Eds.), *Experimenting with uncertainty: Language testing essays in honour of Alan Davies*, pp.138-149. Cambridge, Cambridge University Press.

Turner, C.E. (2002). *Investigating high-stakes test impact at the classroom level.* Presented at 24th Annual Language Testing Colloquium, (LTRC '02). Hong Kong.

Turner, J. (2004) Language as academic purpose, *Journal of English for Academic Purposes, 3*(2), 95–109.

Turner, C.E. (2005). Professionalism and high-stakes tests: Teacher perspectives when dealing with educational change introduced through provincial exams. *TESL Canada Journal*, *23*(2), 54-76.

Turner, C.E. (2008). *The specificity of the "research approach" in classroom studies: Probing the predictability of washback through teacher conceptual and instrumental evidence in Quebec high schools*. Paper presented at the Language Testing and Research Colloquium 2008, Hangzhou.

Turner, C.E. (2009). Examining washback in second language education contexts: A high stakes provincial exam and the teacher factor in classroom practice in

446

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

Quebec secondary schools. *International Journal of Pedagogies and Learning*, *5*(1), 103-123.

Urmston, A. & Fang, F. (2008). *Stakeholder involvement in the language assessment evaluation process: A Hong Kong case study.* Presented at 30th Annual Language Testing Colloquium, (LTRC 08). Hangzhou.

UCLES (2009). University of Cambridge Local Examinations Syndicate. Cambridge.

Valette, R. M. (1994). Teaching, testing, and assessment: conceptualizing the relationship. In C. R. Hancock (Ed.), *Teaching, testing, and assessment*. Lincolnwood: National Textbook Company.

van Lier, L. (1988). *The classroom and the language learner*. Ethnography and second-language classroom research. London: Longman.

Vernon, P.E. (1956). *The measurement of abilities* (2nd ed.). London: University of London Press.

Von Elek, T. (1985). A test of Swedish as a second language: An experiment in self-assessment. In Y. P. Lee, A. C. Y. Fok, R. Lord, & G. Low (Eds.), *New directions in language testing* (pp. 47-55). Oxford: Pergamon Press.

Wall, D. (1996). Introducing new tests into traditional systems: Insights from general education and from innovation theory. *Language Testing, 13*(3), 334-354.

Wall, D. (1998). Impact and washback in language testing. In C. M. Clapham & D. Corson (Eds.) *Language testing and assessment: Encyclopedia of language and education, 7,* 291-302. Dordrecht: Kluwer Academic Publishers.

Wall, D. (1999). *The impact of high-stakes examinations on classroom teaching: A case using insights from testing and innovation theory.* Unpublished doctoral dissertation. England: the University of Lancaster.

Wall, D. (2000). The impact of high-stakes testing on teaching and learning: Can this be predicted or controlled? *System, 28*, 499-509.

Wall, D. (2005). *The impact of high-stakes examinations on classroom teaching*. Cambridge, UK: Cambridge University Press.

Wall, D. & Alderson, C. (1996). Examination washback: the Sri Lankan Impact study. In A., Cumming & R. Berwick (Eds.), *Validation in language testing* (pp. 194-221). Philadelphia: Multilingual Matters

Wall, D., & Alderson, J. C. (1993). Examining washback: The Sri Lankan impact study. *Language Testing*, *10*(1), 41-69.

Wall, D., & Horak, T. (2006). *The impact of changes in the TOEFL examination on teaching and learning in Central and Eastern Europe – Phase 1: The Baseline Study.* TOEFL Monograph Series, MS-34. Princeton, NJ: Educational Testing Service.

Wall, D., & Horak, T. (2008). *The impact of changes in the TOEFL examination on teaching and learning in Central and Eastern Europe – Phase 2, Coping with Change.* Princeton, NJ: Educational Testing Service.

Wallace, M. (1991) Training Foreign Language Teachers: A reflective approach. Cambridge: CUP.

Wallace, M. (1998) Action Research for Language Teachers. Cambridge: CUP.

Wang, H. (2006). *An implementation study of the English as a foreign language curriculum policies in the Chinese tertiary context.* Ph.D. Thesis. Canada: Queen's University, Canada.

Wang, J. (2008). *Identifying the main impact on teacher classroom behaviour: Examining the effects of the CET, teacher beliefs, teacher knowledge bases and teacher proficiency*. Paper presented at the Language Testing and Research Colloquium 2008, Hangzhou.

Wang, J. (2010). *A study of the role of the 'teacher factor' in washback*. Ph.D. Thesis. Canada: McGill University, Canada.

Watanabe, Y. (1992). Washback effects of college entrance examination on language learning strategies. *Japanese Association of College English Teachers, 23,*175 – 194.

Watanabe, Y. (1996). Does grammar translation come from the entrance examination? Preliminary findings from classroom-based research. *Language Testing, 13*(3*)*, 318-333.

Watanabe, Y. (1996b). *Investigating washback in Japanese EFL classrooms: problems of methodology*, *ED* 208-239.

Watanabe, Y. (2004a). *Methodology in washback studies*. Mahwah, NJ: Lawrence Erlbaum Associates.

Watanabe, Y. (2004b). Teacher factors mediating washback. In L.Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods*, pp.129-146. Mahwah, NJ: Lawrence Erlbaum Associates.

Wesche, M. B. (1983). Communicative language testing in a second language. *Modern Language Journal, 67*(1), 41-55.

Weir,C.J.(1990) *Communicative Language Testing*. New York. Prentice Hall.

Wesche, M. B. (1987). Second language performance testing: The Ontario Test of ESL as an example. *Language Testing, 4*(1), 28-47.

Wesdorp, H. (1982). *Backwash effects of language testing in primary and secondary education.* Stichtin Centrum voor Onderwijsonderzoek van de Universite it van Amsterdam. Amsterdam.

Wigglesworth, G. (1994). Patterns of rater behaviour in the assessment of an oral interaction test. *Australian Review of Applied Linguistics, 17*(2), 77–103.

Wigglesworth, G. (1997). An investigation of planning time and proficiency level on oral test discourse. *Language Testing, 14,* 85-106.

Wigglesworth, G. (2001). Influences on performance in task-based oral assessments. In Bygate, M., Skehan, P. & M. Swain: *Task based learning*. Addison Wesley Longman

Wigglesworth, G., Elder, C. (1996). *The Language Testing Cycle: From Inception to Washback*. (Series S Number 13 ed.). Clayton, Australia: Applied Linguistics Association of Australia.

Wiggins, G. (1997). Practicing what we preach in designing authentic assessments. *Educational Leadership*, *54*(4), 18-26. Retrieved June 17, 2010, from EBSCOhost database.

Williams D. (1983). Developing criteria for textbook evaluation. *ELT Journal,* 37(3), 251-255.

Williams, M. & Burden, R. (1997). *Psychology for Language Teachers*. New York: Cambridge University Press.

Willis, J. & Willis, D. (1996) Challenge and Change in Language Teaching. Oxford: Heinemann.

Wilson, B.L. & Corbett, H.D. (1991). *Two state minimum competency testing programs and their effects on curriculum and instruction*, *ED*, 377-251.

Woods, D. (1996). *Teacher Cognition in Language Teaching*. Cambridge: Cambridge University Press.

Woods, D. (2003). The social construction of beliefs in the language classroom. In Kalaja, P., & Barcelos, A. M. F. (Eds.), *Beliefs about SLA: New Research Approaches*. Dordrecht: Kluwer Academic Publishers.

Woods, D. (2006) Who Does What in The 'Management of Language Learning'? Planning and the Social Construction of the 'Motivation to Notice'. In Gieve, S.

and I. Miller (eds). *Understanding Language Classroom.* London: Palgrave Macmillan.

Wong, R.Y.L & James, J.E. (2000). Chapter 10: Malaysia. In H. W. Kam and R. Y. L. Wong (Eds.). *Language Policies and Language Education: The Impact in East Asian Countries in the Next Decade* (p. 209 – 240). Singapore: Times Academic Press.

Xia, J. (2002). A report on investigating the College English teachers' knowledge of and beliefs in foreign language education and their roles in teaching, researching and professional development. *Foreign Language World, 5,* 35-41.

Xia, J. (2005). Interpretation of College English Curriculum Requirements (For Trial Implementation). *Foreign Language World*, *2,* 12-14.

Yan, Z. Y. (1997). *On the harm of examination-oriented education*. Shanxi Education, *5,* 11.

Yang, H. (1999). Language Testing and Language Teaching. *Foreign Language World*, *73*(1), 16-25.

Yang, H. (2000). A look at the actual language competence of Chinese students through the CET 4/6. *Foreign Language World*, 77(1), 46-52.

Yang, H. (2003). The 15 years of the CET and its impact on teaching. *Journal of Foreign Languages, 145*(3), 21-29.

Yang, H. & Jin, Y. (2001). An interpretation of the CET 4/6 scores. *Foreign Language World*, *81*(1), 62-68.

Yang, H. and Weir, C.J. (1998) *The Validity of the College English Test,* Shanghai Foreign Language Education Press.

Yates & Muchisky. (2003) On Reconceptualizing Teacher Education. *TESOL Quarterly*, *37*(1), 135-147.

Ye, J (1998, a). An investigation and reflection on the washback effects of the CET. *Journal of Shanghai Jiaotong University*, 12 (2), 132-136.

Ye, J (1998, b). An investigation and reflection on the washback effects of the CET. *Foreign Language World*, 71 (3), 40-43. Ye, J (1998, a). An investigation and reflection on the washback effects of the CET. *Journal of Shanghai Jiaotong University*, *12* (2), 132-136.

Ying, Y. (2005). Washback effects of the spoken English test on textbook content. *University of China publications, 2005(*1), 11-39.

450

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

Appendix- 1A

Student Questionnaire

Department of English
Jahangirnagar University, Dhaka

The Questionnaire has been developed for the purpose of a research project in the Department of English at Jahangirnagar University, Savar, Dhaka. The questions here are related to the "Washback of the Public Examination on Teaching and Learning English as a Foreign Language (EFL) at the Higher Secondary Level in Bangladesh". The researcher gives you full assurance that the information will be used only for the research purpose, and will be strictly kept confidential.

Thank you for cooperation!

Put a tick mark (√) in the box next to each item, which best expresses your opinion:

	Key: Strongly Agree (SA)= 5; Agree (A)=4; No Opinion (N)=3; Disagree (D)=2; Strongly Disagree(SD)=1					
A	**Curriculum and Syllabus**	SA	A	N	D	SD
		5	4	3	2	1
1	I am aware of the objectives of the HSC syllabus and curriculum.					
2	The present HSC syllabus and curriculum help me learn English.					
3	The teacher teaches us every section in the textbook although some sections are unlikely to be tested in the exam.					
4	I do not care about the syllabus and curriculum while preparing for the examination.					
5	I feel pressure to cover the syllabus before the examination.					

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

451

6	The HSC examination tests my overall competence in English (listening, speaking, reading, and writing)					
7	The teacher gives little attention to practicing the test items necessary for the examination.					
B	**Textbooks and Materials**					
8	The teacher does not tell us the lesson's objectives while teaching.					
9	I think the textbook covers sufficient exercises and opportunities for practicing EFL.					
10	I do not seriously study the textbook materials.					
11	The teacher skips certain sections in the textbook because they are less likely to be tested in the examination.					
12	The textbook, *English for Today* (for classes 11-12) is well-suited to developing communicative competence in English language.					
13	I rely on the test-related materials, such as test papers, past questions, and model questions to take preparation for the examination.					
14	I find interest in studying the textbook materials.					
15	If we study the whole book (*English for Today*), we will perform badly in the final examination.					
16	The teacher uses newspaper articles, radio and television news bulletin, texts of real life incidents, etc. for teaching us English language.					
17	The teacher does not use any modern equipment to teach the English language skills.					
C	**Teaching Methods**					
18	The teacher considers whether we can understand and follow his instruction.					
19	The teacher teaches and explains the text in					

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

452

	English.					
20	The teacher encourages us to ask questions during the class.					
21	The teacher encourages us to speak English in the class.					
22	The teacher uses Bengali along with English to make us understand better.					
23	The teacher teaches whatever he thinks important to teach, no matter whether it is important or not for the examination.					
24	The teacher teaches us the meaning and theme of the topic by explaining the texts line by line, and giving examples.					
25	The teacher does not make us practise how to learn and speak English language but makes us practise how to answer questions in the examination.					
26	My examination results will indicate my language competence and proficiency.					
D	**Classroom Tasks and Activities**					
27	We ignore the task and activities that are not directly related to passing the examination.					
28	The teacher gives us model tests before the final examination starts.					
29	We spend more time practicing grammar because grammar is more likely to be tested in the exam.					
30	My teacher makes us practice and solve the questions of the past examinations.					
31	We give little concentration on learning English language due to examination pressure.					
32	My teacher gives us guidelines on how to answer the questions in the examination.					

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

453

E	EFL Skills and Elements					
33	We practise the EFL skills and elements as per the teacher's design and decision.					
34	Listening is practiced in the class.					
35	Speaking is practiced in the class.					
36	Reading is practiced in the class.					
37	Writing is practiced in the class.					
F	Beliefs, Attitudes, and Perception Related to the Test and Teaching					
38	My parents pressure me to make good results in the examination.					
39	Learning English is more difficult than obtaining good grades in the examination.					
40	I get feedback on my learning from the examination results.					
41	I feel tension for the test preparation.					
42	The present examination system helps me improve language proficiency.					
43	I could learn English better if there were no pressure for good results in the examination.					
44	The results of my HSC Examination will influence my future career.					
45	I may be frustrated if I fail or perform badly in the examination.					

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

454

Appendix-1B

Teacher Questionnaire

Department of English
Jahangirnagar University, Dhaka

The Questionnaire has been developed for the purpose of a research project in the Department of English at Jahangirnagar University, Savar, Dhaka. The questions here are related to the "Washback of the Public Examination on Teaching and Learning English as a Foreign Language (EFL) at the Higher Secondary Level in Bangladesh". The researcher gives you full assurance that the information will be used only for the research purpose, and will be strictly kept confidential.

Thank you for cooperation!

Put a tick mark (√) in the box next to each item, which best expresses your opinion:

	Key= Strongly Agree (SA)= 5; Agree (A)=4; No Opinion (N)=3; Disagree (D)=2; Strongly Disagree(SD)=1					
A	**Curriculum and Syllabus**	SA	A	N	D	SD
		5	4	3	2	1
1	I am aware of the objectives of the HSC syllabus and curriculum.					
2	The present HSC syllabus and curriculum can enhance EFL teaching and learning.					
3	I teach every section in the textbook (*English for Today* for classes 11-12) although some sections are unlikely to be tested in the examination.					
4	I do not care about the syllabus and curriculum while teaching my students.					
5	I feel pressure to cover the syllabus before the final examination.					

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

455

6	The HSC examination tests the overall competence of my students in English (listening, speaking, reading, and writing).					
7	I give more attention to teaching to the syllabus opposed to practicing the test items.).					
B	**Textbooks and Materials**					
8	I follow and communicate the lesson objectives with the students while planning my lessons.					
9	The textbook (*English for Today*) covers exercises that the curriculum has claimed.					
10	I think my students do not seriously study the textbook materials.					
11	I skip certain topics in the textbook because they are less likely to be tested in the examination.					
12	My textbook (*English for Today*) is well-suited to developing the students' communicative competence.					
13	I rely on test papers and sample questions to prepare my students for the examination.					
14	I find interest in teaching textbook materials.					
15	If I teach the whole textbook (*English for Today*), my students will perform badly in the HSC examination.					
16	I use authentic materials along with the textbook for the students' practice of English language in the class.					
17	I use modern aids and equipment to teach and practice English language skills.					
C	**Teaching Methods**					
18	I teach in a way that my students understand everything.					
19	I teach and explain the text in English.					

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

456

20	I encourage my students to ask questions during the class.					
21	I encourage my students to speak English in the class.					
22	I use Bangla along with English to make my students understand better.					
23	I teach whatever I think important to teach, no matter whether it is important or not for the exam.					
24	My role as an English teacher is to transmit knowledge to my students through explaining texts and giving examples.					
25	I do not make my students practise how to learn and speak English language but make them practice on how to answer questions in the examination.					
26	I believe that the test score in the HSC examination in English is an appropriate indicator of a student's English ability.					
D	**Classroom Tasks and Activities**					
27	I ignore the task and activities that are not directly related to passing the examination.					
28	I give model tests to the students to do better in the final examination.					
29	I spend more time teaching grammar because I think grammar is more likely to be tested in the examination.					
30	I make my students practise and solve the questions of the past examinations.					
31	The examination discourages me to teach English language.					
32	I teach test-taking strategies, especially when the examination date gets closer.					

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

457

E	EFL skills and Elements					
33	The examination influences my decision on which language skills are more important to be taught.					
34	Listening is practiced in the class.					
35	Speaking is practiced in the class.					
36	Reading is practiced in the class.					
37	Writing is practiced in the class.					
F	Beliefs, Attitudes, and Perception Related to the Test and Teaching					
38	I feel pressure from my authority to improve my students' test score.					
39	My students can score good marks without improving their English language proficiency.					
40	I get feedback on my teaching from the examination results.					
41	My students suffer from anxiety and tension for the examination.					
42	The present examination system helps my students improve language proficiency.					
43	I could teach English better if there were no pressures for good results in the examination.					
44	Examinations influence my students' future career.					
45	I feel embarrassed if my students fail or perform badly in the examination.					

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

458

COLT Part-A

				DESCRIPTION OF ACTIVITIES AND EPISODES	PARTICIPANT ORGANISATION							CONTENT													CONTENT CONTROL			STUDENT MODALITY					MATERIALS							
					Class			Group		Indiv.		Manag.		Language											Other										Type			Source			
																																		Text						
Start Time	Duration	Material Code	Activities/ Episodes		T ⇄ S/C	S ⇄ S/C	Choral	Same task	Different tasks	Same task	Different tasks	Procedure	Discipline	Form - Vocabulary	Form - Pronunciation	Form - Grammar	Form - Spelling	Function	Discourse	Sociolog.	Narrow	Br	Teacher/Text	Teacher/Text/Stud.	Student	Listening	Speaking	Reading	Writing	Other	MI	Ex	Audio	Visual	L2-NNS	L2-NS	L2-NSA	Student - made
1a	1b	2a	2b		3	4	5	6	7	8	9	10	11	12a	12b	12c	12d	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33

Classroom Observation Schedule – HSC EFL Classroom

Appendix -2B

Modified version of University of Cambridge Observation Scheme (UCOS)

Start Time	Duration	Materials	Activities/ Episodes	DESCRIPTION OF ACTIVITIES AND EPISODES	Skill (W,R,L,S)	Strategy Description	Exam Related	TextTypes	Topic	Text Source	G&V	Reading	Writing	Listening	Speaking
1a	1b	2a	2b		F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12

Appendix -2B

Modified version of University of Cambridge Observation Scheme (UCOS)

Start Time: Class/Observation	Class Duration	Materials	Class Activities/ Episodes	Description of Activities and Episodes	English Language Skills (L, S, R, W)	Use of Bengali	Strategy Description	Examination Related	Text Types	Topic	Text Sources	G & V	Listening	Speaking	Reading	Writing
1a	1b	2a	2b		F2	F3	F4	F5	F6	F7	F8	F10	F11	F12	F13	F14

Appendix- 2C

Self -Made Classroom Observation Checklist (Further Analysis) Part-A

Start Time	Duration	Materials	Activities Episodes	Description of Activities and Episodes	HSC examination			Homework	Interaction	Teacher gives answer	Teacher explain answer	Incidence of laughter
					Exam Information	Exam strategy	Exam Tasks					
1	2	3	4	5	6	7	8	9	10	11	12	13

Self -Made Classroom Observation Checklist (Further Analysis) Part-B

Personality & Professionalism	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
Friendly (Y/N/M)										
Angry (Y/N/M)										
Introvert (Y/N/M)										
Extrovert (Y/N/M)										
Laughter (Y/N/M)										
Shouting (Y/N/M)										
Smiling (Y/N/M)										
Well-behaved (Y/N/M)										
Encouraging (Y/N/M)										
Sincere & Caring (Y/N/M)										
Punctual (Y/N/M)										
Fluent (Y/N/M) in English										
Knowledge of Communicative Competence (E/G/M/P)										
Target language use (A/F/S/N)										
Presentation(E/G/M/P)										
Pronunciation (E/G/M/P)										
Preparedness for Teaching (E/G/M/P)										
Hesitant (Y/N/M)										
Curriculum Knowledge (E/G/M/P)										

Appendix-3A

The Syllabus and Curriculum Analysis Checklist

1	Did the HSC English syllabus and curriculum emphasise the learners' needs?
2	Did the HSC English syllabus and curriculum set goals and objectives?
3	Did the curriculum advocate teaching methods and approaches to be used in the EFL class?
4	Did it put emphasis on communicative competence?
5	Did it suggest sufficient tasks and activities for enhancing communicative competence?
6	How did the present HSC English curriculum treat EFL testing?
7	What procedures were used to determine the contents of a language programme?
8	How could learners' needs be determined?
9	Did it monitor understanding of spoken language by asking questions, seeking clarification, etc.?
10	Was there any scope of reading for literary experience?
11	Was there any scope of reading to be informed of current events such as technology, descoveries, etc?
12	Did it offer and display expanding vocabulary items?
13	Is it possible and feasible to turn EFL classrooms into whole-person events, where body and soul, intellect and feeling, head, hand and heart converge in action?
14	Could EFL learning be a satisfying activity in itself, in the here and now of the classroom?
15	What adventures and challenges are possible under the very conditions of English learning?
16	What was suggested to be done to create situations and scenarios where communication in the target language could be more meaningful?
17	What are the roles of teacher, learners, topic and input in such scenarios in the present English syllabus and curriculum?
18	Can cultural awareness be taught under the present syllabus and curriculum?

Appendix-3B

Textbook Analysis Checklist
***English for Today* for classes 11-12**

1.	Does the textbook correspond to the HSC EFL syllabus and curriculum?
2.	Does the textbook contain exercises for practising the all skills of EFL?
3.	Do the exercises encourage group work, pair work, and individual work?
4.	Are the topics and tasks interesting?
5.	Are the lessons relevant to day to day activities?
6.	Are the instructions clear?
7.	Is the presentation stereotyped and activity boring?
8.	Are the activities student- centred rather than teacher oriented?
9.	Is the textbook appropriate for the HSC level of students and the context?
10.	Is traditional grammar teaching is avoided?
11	Are the grammatical rules presented in a logical manner and in an increasing order of difficulty?
12	Are there enough guidelines and hints for the student?
13	Is it possible for the students to use the textbook on their own?
14	Is the objective expressed in each lesson or unit?
15	Does the title of each lesson indicate the aim of the lesson?
16	Is the sequence of the sections, units, or lessons outlined?
17	Are the new vocabulary words introduced in a variety of ways?
18	Are the new vocabulary words repeated in sequential lessons in order to facilitate the reinforcement of their meaning and use?
19	Are there, in the lesson, exercises that develop comprehension and test the student's knowledge of main ideas, details, and the sequence of ideas?
20	Do the exercises develop meaningful communication by referring to realistic activities and situations?
21	Are appropriate visual materials available?
22	Are the illustrations clear, simple, and free of unnecessary details that may confuse the learners?
23	Are the illustrations printed close enough to the text and directly related to the

	content, to help the learner understand the printed text?
24	Does the lesson mention any topic?
25	Does the lesson provide authentic printed texts (newspapers/ magazine articles, ads, poems, short stories, etc) that have engaged content and tasks?
26	Is the textbook's content obviously realistic, i.e. taken from L1 material not initially intended for ELT purposes?
27	Are different religious and social environments represented?
28	Does the author/s share the linguistic background of its learners?
29	Is the new knowledge connected to the previous lessons?
30	Does the lesson indicate what students should know and be able to do?
31	Does the lesson engage students to use the target language to acquire new information on topics of interest?
32	Does the lesson describe sequence of instructional activities, and assessment?
33	Are teachers given techniques for activating students' background knowledge before reading the text?
34	Does the lesson provide teacher with means for involving students to exercise, practice, test, and review vocabulary words, and grammar?
35	Has the first language a role in teaching foreign language?
36	Is the content appropriate for the students to learn?
37	Are the instructions and examples adequately explained and illustrated for the students?
38	Is it possible to set up groupings varied in response to the nature of learning, i.e., small heterogeneous groups for interaction or cooperative learning, and homogeneous groups for working on achievement or interest?
39	Are the exercises are relevant to every day activities?
40	Are the learning opportunities real, and rich in ways that promote students' engagement and interest?
41	Do the activities enable students to apply their knowledge to practical and real-world situations?
42	Do the activities enable students to be creative?
43	Does the book contain lessons on modern means of communication and innovation of science?
44	Does the book contain lessons especially for exam practice

45	Is the lesson conducive to the students' self-assessment?
46	Are the factors of performance, task and project accessible through the lesson?
47	Can the textbook be termed as well-suited for learning English as a foreign language?
48	Has the textbook given any guidelines for the preparation of the HSC examination?
49	Does the textbook provide any tips and advice for the students?
50	Should the contents of the book be changed now?

Appendix-3E

Test Evaluation Principles and Guidelines

A. Validity and Reliability

Validity: Validity of a test score interpretation can be used as part of the TFF when the following evidence is collected.

- *Criterion-related validity evidence-* this type of evidence (sometimes described as *criterion validity*) refers to whether the test scores under consideration meet criterion variables such as school or college grades and on the job-ratings or some other relevant variable.

- *Concurrent validity-* the test scores correlate with a recognized external criterion which measures the same knowledge or ability.
- *Construct validity-* the extent to which a test measures the construct it intends to measure.
- *Content validity-* the extent to which test content represents the course of study.
- *Face validity-* the extent to which users consider a test to be an acceptable measure of the ability they wish to measure.
- *Predictive validity-* the extent to which tests result be considered as a measure for real-life or indicator of proficiency in English.

Reliability: This type of evidence refers to the reliability or consistency of test scores in terms of consistency of scores among different testing occasions (describes as *stability* evidence), among two or more different forms of a test (*alternate form* evidence), among two or more raters (*inter-rater* evidence), and in the way test items measuring a construct functions (*internal consistency* evidence).

- *Relevance:* the extent to which it is necessary that students are able to perform the tasks.
  Authenticity: the extent to which the situation and the interaction are meaningful and representative in the world of the individual user.
  Balance: each relevant topic/ability receives an equal amount of attention.

B. Testing of language skills

1. To what extent different skills (e.g. listening, speaking, reading and writing) of English language are tested in the HSC examination?

2. To what extent the linguistic elements (e.g. vocabulary items, grammar, etc) are tested in the HSC Examination?

3. To what extent the inherent knowledge are tested?

4. Is the HSC examination in English a reliable test?

5. Is the HSC examination in English a valid test?

6. To what extent the testee was able to complete his/her role in the communicative task assigned?

C. Principle

The following general principles are maintained:

- *Principle 1*: A test ought to be fair to all test takers, that is, there is a presumption of treating every person with equal respect.
- *Principle 2*: A test ought to have comparable construct validity in terms of its test-score interpretation for all test takers.
- *Principle 3:* A test ought not to be biased against any test taker groups, in particular by assessing construct-irrelevant matters.
- *Principle 4*: A test ought to bring about good in society, that is, it should not be harmful or detrimental to society.
- *Principle 5:* A test ought to promote good in society by providing test score information and social impacts that are beneficial to society.
- *Principle 6*: A test ought not to inflict harm by providing test-score information or social impacts that are inaccurate or misleading.

D. Specifications and Guidelines
The following specification of Bachman and Palmer (1996) format was maintained:

- *Purpose*: an explicit statement of how the test item/task should be used.
- *Definition of the construct*: a detailed description of the construct, or particular aspect of language ability, that is being tested. This includes the

inferences that can be made from the test scores, which overlaps with the purpose of the test.

- *Setting*: a listing of the characteristics (physical location, participants and time of administration) for the setting in which the test will take place.

- *Time allotment:* the amount of time allowed for completing a particular set of items or a task on the test.

- *Test rubric:* the test rubric facets are portrayed by Bachman and Palmer (1996) as those characteristics of the test that provide the structure for a particular test. These characteristics include: instructions and the structure of the test (e.g., how the test is organized), time allotment (e.g., the duration of the test as a whole and of the individual tasks), and scoring procedures (e.g., how the language that is used will be evaluated, or scored).

- *Input:* the input facets examine the format in which input is presented and the characteristics of the language that are embodied in the input. The format includes features such as channel, form, language, length, type, degree of 'speededness', and vehicle of input delivery, while the language characteristics include both organizational and pragmatic characteristics of how the language is organised.

- *Instructions*: a listing of the language to be used in the directions to the test takers for that particular item/task.

- *Characteristics of the input and expected response*: essentially a description of what will be presented to the test take and what they will be expected to do with it.

- *Scoring methods:* a description of how the test taker response will be evaluated.

E. Access

Access of a test can be used as part of the test analysis when the following evidence is collected.

- *Educational access:* This refers to whether a test is accessible to test takers in terms of *opportunity to learn* the content and to become familiar with the types of tasks and cognitive demands.

- *Financial access:* This refers to whether a test is financially *affordable* to test takers.

- *Geographical access:* This refers to whether a test site is accessible in terms of distance to test takers.

- *Personal* access: here refers to whether a test offers certified test takers with physical and learning disabilities with appropriate test accommodations. The 1999 *Standards* and the *Code* (1988) calls for accommodation in order that test takers who are disabled are not denied access to tests that can be offered without compromising the construct being measured.

- *Conditions or equipment access:* This refers to whether test takers are familiar with to test taking equipment (such as computers), procedures (such as reading a map) and conditions (such as using planning time).

F. Social consequences

The social consequences of a test can be used as part of the test fairness framework when evidence regarding the following need to be collected:

- *Washback:* This refers to the effect of a test on instructional practices, such as teaching, materials, learning, test taking strategies, etc.

- *Remedies:* This refers to remedies offered to test takers to reverse the detrimental consequences of a test such as re-scoring and re-evaluation of test responses, and legal remedies for high-stakes tests. The key fairness questions here are whether the social consequences of a test and/or the testing practices are able to contribute to societal equity or not and whether there are any pernicious effects due to a particular test or testing programme.

Appendix-3E

Test Evaluation Principles and Guidelines

A. Validity and Reliability

Validity: Validity of a test score interpretation can be used as part of the TFF when the following evidence is collected.

- *Criterion-related validity evidence-* this type of evidence (sometimes described as *criterion validity*) refers to whether the test scores under consideration meet criterion variables such as school or college grades and on the job-ratings or some other relevant variable.

- *Concurrent validity-* the test scores correlate with a recognized external criterion which measures the same knowledge or ability.
- *Construct validity-*the extent to which a test measures the construct it intends to measure.
- *Content validity-* the extent to which test content represents the course of study.
- *Face validity-* the extent to which users consider a test to be an acceptable measure of the ability they wish to measure.
- *Predictive validity-*the extent to which tests result be considered as a measure for real-life or indicator of proficiency in English.

Reliability: This type of evidence refers to the reliability or consistency of test scores in terms of consistency of scores among different testing occasions (describes as *stability* evidence), among two or more different forms of a test (*alternate form* evidence), among two or more raters (*inter-rater* evidence), and in the way test items measuring a construct functions (*internal consistency* evidence).

- *Relevance:* the extent to which it is necessary that students are able to perform the tasks.
  Authenticity: the extent to which the situation and the interaction are meaningful and representative in the world of the individual user.
  Balance: each relevant topic/ability receives an equal amount of attention.

B. Testing of language skills

1. To what extent different skills (e.g. listening, speaking, reading and writing) of English language are tested in the HSC examination?

2. To what extent the linguistic elements (e.g. vocabulary items, grammar, etc) are tested in the HSC Examination?

3. To what extent the inherent knowledge are tested?

4. Is the HSC examination in English a reliable test?

5. Is the HSC examination in English a valid test?

6. To what extent the testee was able to complete his/her role in the communicative task assigned?

C. Principle

The following general principles are maintained:

- *Principle 1*: A test ought to be fair to all test takers, that is, there is a presumption of treating every person with equal respect.

- *Principle 2*: A test ought to have comparable construct validity in terms of its test-score interpretation for all test takers.

- *Principle 3:* A test ought not to be biased against any test taker groups, in particular by assessing construct-irrelevant matters.

- *Principle 4*: A test ought to bring about good in society, that is, it should not be harmful or detrimental to society.

- *Principle 5:* A test ought to promote good in society by providing test score information and social impacts that are beneficial to society.

- *Principle 6:* A test ought not to inflict harm by providing test-score information or social impacts that are inaccurate or misleading.

D. Specifications and Guidelines

The following specification of Bachman and Palmer (1996) format was maintained:

- *Purpose*: an explicit statement of how the test item/task should be used.

- *Definition of the construct*: a detailed description of the construct, or particular aspect of language ability, that is being tested. This includes the

471

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

inferences that can be made from the test scores, which overlaps with the purpose of the test.

- *Setting*: a listing of the characteristics (physical location, participants and time of administration) for the setting in which the test will take place.

- *Time allotment:* the amount of time allowed for completing a particular set of items or a task on the test.

- *Test rubric:* the test rubric facets are portrayed by Bachman and Palmer (1996) as those characteristics of the test that provide the structure for a particular test. These characteristics include: instructions and the structure of the test (e.g., how the test is organized), time allotment (e.g., the duration of the test as a whole and of the individual tasks), and scoring procedures (e.g., how the language that is used will be evaluated, or scored).

- *Input:* the input facets examine the format in which input is presented and the characteristics of the language that are embodied in the input. The format includes features such as channel, form, language, length, type, degree of 'speededness', and vehicle of input delivery, while the language characteristics include both organizational and pragmatic characteristics of how the language is organised.

- *Instructions*: a listing of the language to be used in the directions to the test takers for that particular item/task.

- *Characteristics of the input and expected response*: essentially a description of what will be presented to the test take and what they will be expected to do with it.

- *Scoring methods:* a description of how the test taker response will be evaluated.

E. Access

Access of a test can be used as part of the test analysis when the following evidence is collected.

- *Educational access:* This refers to whether a test is accessible to test takers in terms of *opportunity to learn* the content and to become familiar with the types of tasks and cognitive demands.

472

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

- *Financial access:* This refers to whether a test is financially *affordable* to test takers.

- *Geographical access:* This refers to whether a test site is accessible in terms of distance to test takers.

- *Personal* **access**: here refers to whether a test offers certified test takers with physical and learning disabilities with appropriate test accommodations. The 1999 *Standards* and the *Code* (1988) calls for accommodation in order that test takers who are disabled are not denied access to tests that can be offered without compromising the construct being measured.

- *Conditions or equipment access:* This refers to whether test takers are familiar with to test taking equipment (such as computers), procedures (such as reading a map) and conditions (such as using planning time).

F. Social consequences

The social consequences of a test can be used as part of the test fairness framework when evidence regarding the following need to be collected:

- *Washback:* This refers to the effect of a test on instructional practices, such as teaching, materials, learning, test taking strategies, etc.

- *Remedies:* This refers to remedies offered to test takers to reverse the detrimental consequences of a test such as re-scoring and re-evaluation of test responses, and legal remedies for high-stakes tests. The key fairness questions here are whether the social consequences of a test and/or the testing practices are able to contribute to societal equity or not and whether there are any pernicious effects due to a particular test or testing programme.

Appendix-3F

Answer Scripts Analysis Guidelines and Checklist

A. Principles and Guidelines for answer scripts evaluation
(Based on Morrow, 1991; Brown, 2003; and Li, 2009)

1.	Evaluation has two components: information of achievement or performance, and a standard that provides a base for measuring.
2.	In a fundamental sense, evaluation means placing a value on some entity thus expressing an indication of its worth.
3.	Answer scripts evaluation can no more be a totally rational process than instruction.
4.	Examiners' sensitivity and values are inevitable factors in the effectiveness of their evaluations.
5.	Evaluation is one of three major components of teaching, along with planning and instruction.
6.	Evaluation of scripts has been described as one of the components of teaching..
7.	The answer scripts evaluation process requires deliberation in a number of important areas.
8.	The evaluation process also typically considers a range of general areas relating to task performance: level of difficulty, task clarity, timing, layout, degree of authenticity, amount of information provided, and familiarity with the task format.
9.	Communicative language tests can be evaluated in terms of their performance within specific learning contexts.
10.	The evaluation process involves analyzing test results in light of both test specifications and program objectives.

B. Analysis Checklist

1.	Who are involved in the scoring/marking of the HSC EFL answer scripts?
2.	Are there any guidelines for scoring the answer scripts in English?
3.	Is there any arrangement of training for examiners to examine and marking the scripts?
4.	Are the examining and marking procedures reliable?
5.	Is the answer scripts evaluation valid?
6.	How long does an examiner take to examine a script?
7.	Which skills and linguistic elements are examined comfortably?
8.	Which are the maximally attempted questions?
9.	Which are the high scoring items/questions?
10.	Which items/questions are answered first by the students?
11.	What is the highest score of the students?
12.	What is the lowest score of the students?
13.	To what extent does the score vary from examiner to examiner?
14.	Do the scores of the testees reflect their performance in English?
15.	Are the examiners biased?

Appendix- 4A

Interview Questions for EFL Teachers

	Before and during the classroom observation:
1.	Do you tell me something about your language learning experience?
2.	Do you say something about yourself, such as your educational background and teaching experience?
	EFL Teacher Interview Questions (After the classroom observation):
3.	Would you please describe your experience following the HSC English Syllabus?
4.	What do you think of the clarity of the national HSC Syllabus? Is it clear and easy to follow? If not, why not?
5.	What do you think of the practicality of the syllabus in terms of your language teaching? Is it practical for you to use? If not, why not?
6.	How appropriate is the HSC English syllabus and curriculum for your teaching context?
7.	To what extent do you think the textbooks accurately represent the HSC English Syllabus and curriculum? If not, where are they lacking?
8.	What is the benefit of the College English Test from your perspective as a teacher?
9.	What impact, if any, does the College English test have on your classroom teaching?
10.	What impact, if any, does the College English test have on your students' learning?
11.	How do you actually conduct your classroom teaching to achieve the objectives stated in the official syllabus?
12.	How much support do you receive from your department head regarding how you teach College English?
13.	What kinds of support would you like to receive from your department head in your teaching of College English?
14.	What are the most challenging aspects that you face regarding implementing the College English Curriculum?
15.	What is the rationale for you to conduct your class in English / in Bengali?
16.	Why do you organize your classroom activities that allow your students to work with their peers / to work alone?

Appendix- 4B

Interview Questions for Examiners

1.	Can you describe your background as an English examiner?
2.	How long have you been an examiner of English?
3.	Do you enjoy examining the scripts?
4.	Have you received any training for examining the scripts?
5.	Are you provided with any written guidelines on evaluation?
6.	Are you influenced by any interference and external pressure?
7.	Do you believe that the HSC examination reflects the goals and objectives of the HSC English curriculum?
8.	How long do you take to examine a script?
9.	Why do you take so little/much time?
10.	Do you think that the present EFL test is a valid test?
11.	Do you think the HSC examination in English is a reliable test?
12.	Do you think that the test scores of the HSC examination are an appropriate indicator of a students' proficiency in English?
13.	What difficulties do you face while examining the scripts?
14.	Do you think that the present EFL test helps enhance EFL teaching and learning?
15.	Do you think learning a language is to accumulate the knowledge of grammar, vocabulary, structures and rules?
16.	What are the major shortcomings of the present EFL test in terms of items and contents?
17.	Do you think that the present examination should be modified or revised?
18.	What are your suggestions for the improvement of the test?

477

Language in India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D.   <Pages 288-786>

Appendix- 4C

Interview Questions for Curriculum Specialists

1.	Do you think the present HSC English syllabus is well-suited to developing communicative competence in English language?
2.	What was the rationale for changing/modification of the old HSC English syllabus of 1990 version?
3.	What changes have been made to the syllabus since then?
4.	How appropriate is the HSC English syllabus for the Bangladesh teaching context?
5.	How practical is the HSC English syllabus for the HSC English language teaching and learning?
6.	To what extent does the NCTB expect the colleges to follow the syllabus?
7.	How is the syllabus communicated to future education?
8.	Has the HSC examination in English a positive influence on teaching (e.g., they help focus teaching, provide feedback on teaching
9.	To what extent does the NCTB intend to get teachers involved in the development/modification of the syllabus?
10.	To what extent does the authority take into account teachers' language proficiency when delivering the syllabus?
11.	How is the intended curriculum interpreted by the classroom teachers?
12.	Do you think the present HSC English syllabus and curriculum developed in 2000 requires modification?
13.	What is your opinion on the present examination system?
14.	Do you think the results of the HSC examination in English indicate reasonable level of competence in English of the students?
15.	Do you think the HSC examination is a valid and reliable test?