

Full Length Research Paper

Investigating the washback effect of the Pakistani Intermediate English Examination

Asma Aftab, Sabeen Qureshi and Isabel William

Kinnaird College for Women, Lahore, Pakistan.

Received 6 November, 2013; Accepted 30 June, 2014

The impact of a test on teaching and learning is commonly referred to as the washback effect. This study investigated the nature and scope of the washback effect from the Intermediate English examination on teachers and students of a public sector college in Pakistan. The research relied on qualitative approach utilizing interviews to collect data from six teachers and six students. The data was analysed using open-coding. The results revealed that there seems to be a strong negative washback from the examinations on teaching methodology, content and learning. The pedagogical implications of the current study in the form of recommendations related to English language assessment procedures include the employment of formative assessment, the use of authentic tasks and focus on all four language skills.

Key words: Washback effect, English language examination, language assessment.

INTRODUCTION

Language teaching and language testing are closely related. This correlation is a central issue in the case of English language learners and it is impossible to analyse either of them without referring to the other. Language testing can serve to evaluate the achievements of language teaching and, at the same time, it has some washback effects on teaching, which may be conducive to language teaching or may result in negative repercussions. Thus, language testing has a vital role to play in language learning.

According to Hughes (1989), testing is important since it provides information about individual's language ability and their achievement. It is necessary to have a common yardstick (which tests provide) to make meaningful comparisons. Testing is used by and has impact on

schools, colleges and administrators. Assessment is also perceived to have an impact on what and how teachers teach. Linn (1992) assumes that all the above mentioned consequences of testing need to be evaluated. The foremost of these is the identification of the intended effects of assessment on teaching and learning. This effect is known as washback in testing literature. It is argued that the deleterious effects associated with some high-stake tests could be overcome by performance based assessment which could be linked with the goals of instruction so that the instruction and learning could improve (Noble and Smith, 1994). Nowadays, test scores inform and help plan important educational decisions concerning learners, teaching and educational programmes. Examination scores are also used for various

*Corresponding author. E-mail: elt@kinnaird.edu.pk

Authors agree that this article remain permanently open access under the terms of the [Creative Commons Attribution License 4.0 International License](https://creativecommons.org/licenses/by/4.0/)

social purposes, which makes the study of washback a high interest phenomenon with the ability to influence the activities of teachers, policy makers, learners and other stakeholders in the educational contexts.

What is washback?

A phenomenon that was discussed in education but not acknowledged until the early 1990s was the impact of exams on teaching and learning commonly known as washback effect. According to Alderson and Wall (1993, p.115), testing affects teaching and learning and “tests are held to be powerful determiners of what happens in the classroom”. There are numerous definitions of washback in the relevant literature. Bachmann and Palmer (1996, p. 35) state that washback is a more complex phenomenon than simply the influence of a test on teaching and learning. They feel that the impact of a test should be evaluated with reference to the contextual variables of society’s goals and values, the educational system in which the test is used, and the potential outcomes of its use. Bailey (1996) distinguishes between the influences on learners and those on the programme. The latter involves judging students’ language in relation to the expectations of the curriculum in order to determine whether the school as a whole performs well or whether teaching methods and textbooks are effective tools for achieving curricula goals.

Types of washback

Generally washback can be analyzed according to two major types (positive and negative) depending on whether it has a beneficial or harmful impact on educational practices (Hughes, 1989). Bailey (1996) stresses attention to four aspects in order to ensure positive washback effects – the purpose of language learning; authenticity of testing; students’ autonomy and self assessment; and the feedback of test results. 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. For instance, if the writing skill is only assessed by multiple choice items, then “there is great pressure to practice such items rather than to practice the skill of writing itself” (Davies et al., 1999, p. 225). Thus, teachers may teach to the test leading to an undesirable limitation of the course content in order to help the learners achieve high scores (Alderson and Wall, 1993). Wash-back can also be classified into ‘overt’ and ‘covert’. The former is generally negative and involves the explicit use of examination papers or examples from textbooks that emphasize the skills used in examinations. On the other hand, ‘covert’ washback deals with assumptions about how students learn. Cheng et al. (2004) distinguish between ‘strong’

and ‘weak’ washback effects. The former influences everything that happens in the classroom and where all teachers are teaching in the same way. An example of this scenario might be observed in the form of changes in teaching methodology and classroom activities as a consequence of the modifications in the test format from subjective to multiple choice questions. Conversely, ‘weak’ washback affects some teaching practices and is more likely to be related to teacher factors such as attitudes towards the test than the test itself.

Key washback studies

A number of research studies have been conducted to investigate the phenomenon of washback and explore how it affects the various stakeholders. One of the later studies was undertaken by Cheng (1997) in the context of secondary school examinations in Hong Kong. She examined the possible washback effects of the 1994 Revised Hong Kong Certificate of Education Exam in English (HKCEE), which was considered a high-stake exam. She investigated the impact both at the macro-level on the different stakeholders like the Examination Authority, textbook publishers, and the tertiary institutions; and at the micro-level on teachers’ attitudes, teaching content and classroom interactions. Her research was wide-scale; she undertook classroom observations of 45 lessons of 12 teachers, utilized questionnaires for 550 teachers and 1700 students, and conducted interviews of an unspecified number of teachers. Her data revealed a range of attitudes and behavioral changes. The examination was supposed to encourage task-based learning. The teachers felt positively about the examination that enabled the students to use English more practically and authentically. However, the teaching methodology which was teacher-centered remained unchanged. The teaching content focused more on developing listening and speaking skills as was the requirement of the revised examination. These findings showed that while classroom content may change because of a test, the way teachers instruct does not change to any significant degree.

Green (2007) studied the IELTS preparation classes to find out if they helped to improve IELTS writing scores. Participants were asked to take a pre-test and a post-test before and after their courses. Questionnaires were used to collect data about participants regarding their background, motivation, classroom tasks and learning strategy use. The findings highlighted that test driven instruction does not necessarily increase students’ scores implying that ‘teaching to test’ is no more effective in boosting test scores than teaching the targeted skills. Though the concept of washback has been explored in different studies around the world, there is hardly any illuminating research on washback effect of testing in Pakistan, even though in this context education is mostly examination driven (Shamim, 2011). Moreover, the

government has not taken any steps to explore students' learning at the Higher Secondary level. Thus, there is a dire need to study the examination system at the secondary level so that insights could be gained into teaching and learning which could inform decisions regarding curriculum planning, teaching methods, examination content and other contextual factors.

METHODOLOGY

The present research project aims to explore the nature and scope of the washback effect from the Pakistani Intermediate English examination. It also attempts to highlight the teachers' and students' perceptions about the washback effect of this examination. The present research is situated within the constructivist paradigm and uses qualitative data collection methods since, as asserted by Bailey (1996), washback is a real world phenomenon which probably is not accessible to measurement. Constructivists view reality as socially created and aim to understand human experience as was being attempted through this study. The researchers were interested in generating an understanding of the selected context and the role the participants' perceptions play in interpreting washback. The gathered data revealed the respondents' perceptions and helped the researchers in understanding how washback operates and influences on teaching and learning.

The sample consisted of 6 teachers and 6 students of a women's college situated in one of the main urban centers of Pakistan. The instructors had varying experience and were a representative group of the faculty at the college. One of the selected teachers was asked to choose a heterogeneous group of six students for the interviews. All the participants were female; the study did not focus on gender differences. Since the researchers were only interested in studying the washback in one institution, cluster sampling technique was considered appropriate for this study. This research utilized semi-structured interviews as the main tool to collect the data. Traditionally, many washback studies used questionnaires to study washback. However, this tool can prove to be limiting (as far as the relevance of responses is concerned) and relatively less motivating for the respondents. In contrast, the interviews can draw out "deep" information about the participants' opinions and beliefs regarding a topic. Importantly, the semi-structured format can prove to be helpful to the interviewer since this can provide him/her opportunities to explore and probe freely within some identified inquiry areas. The study was limited to one higher education institution in Pakistan because of practical constraints. However, the project managed to highlight a few significant aspects related to washback in Pakistani higher education scenario paving way for more detailed exploration of English assessment in the country. The obtained findings can also help to improve our understanding of this phenomenon in other contexts where English is taught as a second or foreign language.

The Intermediate examination

After the completion of grade 11 and 12, the students are required to take the public Intermediate examinations. The successful candidates are awarded the Higher Secondary School Certificate (HSC). This examination determines whether students will be able to undertake tertiary education or obtain better jobs. The scores obtained in this examination function as selection criteria for universities. The Intermediate examinations are administered by a regional Board of Intermediate and Secondary Education (BISE). The English Intermediate examination comprises of two papers. Each paper carries 100 marks. Students appear for Paper I at the

end of year XI and paper II at the end of year XII. Each paper has a subjective and objective section. The content of the examination, conducted in Lahore by the Lahore Board of Intermediate and Secondary Education, is based heavily on the prescribed textbooks which consist of selections from English literature.

ANALYSIS OF DATA

After interpreting the findings it was identified that there is negative washback from the Intermediate examination at the selected college. The following factors contributed towards the negative influence on teaching and learning within the context of the institution.

1. The results clearly indicated that the teachers are teaching towards the examination and their teaching appears to be directly influenced by the assessment procedures. The teacher's perception of the Intermediate examination as high stake affects their teaching practices as they focus on examination related activities to help students score better, but research suggests that such practices may increase test scores without necessarily increasing understanding (Noble and Smith, 1994).
2. The instructors felt that the examination restricted them to content based teaching. Moreover, the syllabus and the examination did not allow communicative language teaching. Lyons (1998) points out that the tension between pedagogic and ethical decisions occur when teachers believe that "tests run contrary to the principles and practices of current approaches to language learning" (Bailey, 1996, p.259) and when they believe that the most effective way for their students to achieve higher test scores is to be provided practice in examination related tasks.
3. The teachers tend to ignore speaking and listening skills in favour of reading and writing that are tested in the examination. The teachers justify the negligence as they are mainly focused on preparing the students for the examinations.
4. The teacher's responses also revealed that though the examination focused on reading and writing skills, it neglected assessing the higher order cognitive skills. The writing skills are mostly tested through memorized answers; the reading comprehension questions are text based and direct, and do not encourage critical thinking. Such a misrepresentation of test construct affects its validity and causes negative washback (Brown, 2000).
5. The multiple choice (objective) assessment also yields negative washback on the grounds that it restricts test content and encourages poor teaching practices (Hughes, 2003). It has also been assumed that these types of objective tests only assess the lower order skills of recall or recognition. It has made scoring high on the examinations relatively easy for the students.
6. According to Hughes's (1993), the product of the examination is concerned with whether it leads to learning or not. In case of the Intermediate examination,

the students reported that the examination did not help in improving their English language skills. This is because their learning is passive rather than active. Most of the undertaken activities are directly linked to examination questions. As Nunan (1998, p.224) has pointed out, tasks of this kind “directly call on what learners already know about a language”, but do little to extend what they can already do. As acknowledged by the respondents themselves, memorization of answers does not help to improve language abilities.

7. The Intermediate students heavily depend on rote memorization from guides (test practice books) and are of the view the examination itself promotes these practices.

8. There is washback on materials as well since the teachers adopt a textbook oriented methodology and do not leave anything from the textbooks based on the beliefs that the students may be tested on it. The use of past examination papers as practice material also indicates that teachers make use of examination related activities to train the learners for the examination. Fullilove (1992) while discussing teaching materials designed to prepare students for public examinations states that such tests are “little more than cloners of past exam papers” (p.139). The Intermediate students reported relying on the predictable pattern of past papers to assist them in preparing and scoring high marks.

The intensity of washback can be described as strong or weak and could be a function of the importance of test's consequences: that is, the higher the stakes, the stronger the intensity. Hughes (1993, cited in Green, 2007) asserts that intensity varies in relation to participants' perceptions of test stakes and test difficulty. Washback is said to be strongest or most intense where participants:

1. *Value success on the test above developing skills for the target language use domain*
2. *Consider success on the test challenging, but both attainable and amenable to preparation*
3. *Work in a context where these perceptions are shared (or dictated) by other participants*

The results have revealed that both teachers and students consider scoring high in the Intermediate examinations of utmost importance. Both parties believe that it can be attained through proper practice of examination related tasks. Thus, the above mentioned conditions are met by the participants of the study in the light of the importance of the English examination. Through this study it has been established that the Intermediate English examination is perceived to be a high stakes examination. Moreover, in the context of this study, the data has revealed that teachers' practices are influenced by examination content and format which also affects students' learning. A few teachers mentioned conducting creative activities to develop the neglected skills of students, but these claims could not be

investigated since the scope of this project was limited and classroom observations could not be conducted to verify the reliability of the teachers' responses. Thus, we cannot say that the examination determines everything that happens in the classroom.

The washback on learning is strong since students also demand practice of examination related activities and rely on cramming of material from guides (practice books). Their attitude towards the English examination suggests that they treat English as a subject and not a language and are not concerned with acquiring skills to become proficient in the language. The learners themselves confessed that their English proficiency level had not improved during the two years they studied the Intermediate syllabus. On the basis of the respondents' insights into the classroom practices it can be determined that the Intermediate English examination has a strong negative washback on teaching and learning at the college. The washback effect of the Intermediate examination on teaching and learning is represented through the following model of washback (As shown in Figure 1). The figure represents the complex nature of examination washback mediated through the perceptions of the key stakeholders (the teachers and students). In the model, the examination characteristics interact with the perceptions of teachers and students regarding examination requirements, English language learning and use of results to mediate examination washback. The examination preparations mediate between participants' perceptions and washback on teaching, learning and materials.

In the present study the high stakes of the examination influenced the participants' perceptions and behavior. The examination characteristics like format and content also affected the participants' beliefs about examination requirements. This interaction of the examination characteristics with the participants' perceptions influences examination preparation activities. There was a negative washback effect on teaching, learning and materials. The teachers heavily relied on practicing of examination related tasks and focused on covering everything in the textbooks. The students depended on rote learning answers from guides. The examination promoted assessment oriented methodology with focus on only those language skills that are tested formally. Thus, students' language learning needs appear to be neglected.

RECOMMENDATIONS

The following recommendations (based on the findings of this small-scale study) are proposed.

Assessment of learning

The major purpose of assessment is and should be to

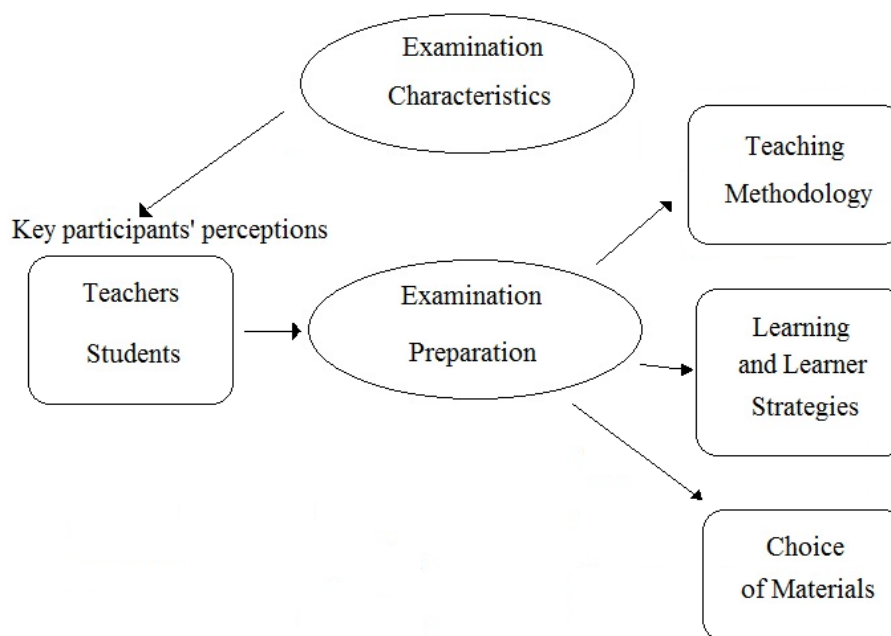


Figure 1. Conceptual model of Washback.

support teaching and learning process (Gipps, 1996). It is through formative assessment that learning can be enhanced and deepened. Real learning only takes place when it enhances students; understanding, enabling them to interpret and apply it in a totally different context than in which it was learnt (Harlen and James, 1997).

According to Black and William (1998), formative assessment helps to modify teaching and learning activities. In the present study, the assessment is summative and the students' grade or rank does not properly convey the extent of their learning, whereas if formative assessment is used the students will get feedback from their teachers which will help them improve.

Performance assessment

Performance assessment based on the constructivist model of learning is defined by Gipps (1996) as "a systematic attempt to measure a learner's ability to use previously acquired knowledge in solving novel problems or completing specific tasks (p.99). The researchers believe that high-stakes testing could drive reform if it followed better psychology and employed more appropriate measurement forms like performance assessment. It is proposed that if better (task-based and integrated) assessment is used it will result in better teaching and learning and thus may result in positive washback.

Examination content

The Intermediate English examination is an achievement test based on detailed syllabus comprising of textbooks.

However, the examination does not assess all the four language skills, but limits itself to reading and writing skills. Thus, it fails to prepare the students for the types of tasks required at University level. In order to avoid this of curricular goals with examination content, an alternative approach could be adopted; the content can be based directly on the objectives.

Variety of assessment tasks

In order to make the assessment process more reliable, the students should be given more opportunities to show the extent of their knowledge through the use of multiple tasks. The intermediate examination makes use of indirect testing to measure the language skills of the candidates (as noted by Raza, 2009, in the context of overall assessment in Pakistan). Hughes (1989) argues that the relationship between performance on indirect tests and performance in the skills under focus is rather weak and uncertain. In contrast, authentic and direct assessment can lead to positive washback (Hughes, 2003).

Comprehensive assessment of language skills

The respondents' perceptions revealed that the Intermediate English examination scores do not reflect the integrated language abilities of the candidates. It is proposed that an oral skills component should be incorporated in the internal college assessment. Since the teaching at the selected college is examination

oriented, such revision in assessment measures will lead to focus being diverted to these neglected skills.

CONCLUSION

This study explored washback from the intermediate examination through the perceptions of teachers and students. There is need for more empirical studies in the Pakistani context to explore Washback from high-stake public examinations like the Intermediate. Studies should use classroom observations to investigate teachers' and students' behaviours. In addition, the present research project utilized a very small sample of respondents; subsequent studies can be more broad-based incorporating a large number of students and teachers. Moreover, since this study just focused on two participants (students and teachers), research on other participants influenced by washback (like the test developers, curriculum planners, materials developers, administrators) (Cheng 1997) can also prove to be illuminating. In fact, the washback phenomenon needs to be studied at the micro and macro level so that better steps could be undertaken to reform the examination system resulting in positive washback. The findings obtained from Pakistani contexts can also be compared to the data derived from washback studies having been conducted in other countries; this can go a long way in improving assessment mechanisms in related contexts.

Conflict of Interests

The author(s) have not declared any conflict of interests.

REFERENCES

- Alderson JC, Wall D (1993). Does Washback Exist? *Applied Linguistics*, 14(2):115-129.
- Bachmann L, Palmer A (1996). *Language Testing in Practice*, Oxford: Oxford University Press. http://www.goodreads.com/book/show/1605674.Language_Testing_in_Practice
- Bailey K (1996). Working for Washback: A Review of the Washback Concept in *Language Testing*. *Language Testing* 13:257-279.
- Black P, William D (1998). Assessment and Classroom Learning. *Assessment in Education: Principles, Policy Pract.* 5(1):7-74.
- Brown JD (2000, January). University Entrance Examinations: Strategies for Creating Positive Washback on English Language Teaching in Japan. *Shiken: JALT Testing & Evaluation SIG Newsletter*, (3):2, 4-8. Retrieved February 27, 2005, from http://jalt.org/test/bro_5.htm
- Cheng LY (1997). How does Washback Influence Teaching? Implications for Hong Kong. *Lang. Educ.* 11(1):38-54.
- Cheng L, Watanabe Y, Curtis A (2004). *Washback in Language Testing: Research Contexts and Methods*, Mahwah, NJ, London: Lawrence Erlbaum.
- Davies A, Brown C, Elder K, Hill T, Lumhey, McNamara (1999). *Dictionary of Language Testing: Studies in Language Testing*, Cambridge: Cambridge University Press. <http://www.amazon.com/Dictionary-Language-Testing-Studies/dp/0521658764>
- Fullilove J (1992). The Tail that Wags. *Institute of Language in Education J.* 9:131-147.
- Gipps C (1996). Assessment for Learning. In: Little A. & Wolf A. (Eds.), *Assessment in Transition*, Oxford: Pergamon.
- Green A (2007). Washback to Learning Outcomes: A Comparative Study of IELTS Preparation and University Pre-session Language Courses. *Assessment Educ.* 14(1):75-97.
- Lyons –Hamp L (1998). Ethics in Language Testing. In: Corson D. and Clapham C. (Eds.) *Language Testing and Assessment: Encyclopedia of Language Educa.* Amsterdam: Kluwer Academic Publisher 7:323-333.
- Harlen W, James M (1997). Assessment and Learning: Differences and Relationships between Formative and Summative Assessment. *Assessment in Education: Principles, Policy Pract.* 4(3):365-379.
- Hughes A (1989). *Testing for Language Teachers*, New York: Cambridge University Press.
- Hughes A (1993). Backwash and TOFEL 2000. Unpublished manuscript, University of Reading.
- Hughes A (2003). *Testing for Language Teachers*. (2nd ed.), Cambridge: Cambridge University Press.
- Linn RL (1992). Achievement Testing. In: M.A.C. Alkin (ed). *Encyclopedia of Educational Research*, Vol I, (6th ed.), pp.1-12, New York: Macmillan.
- Noble AJ, Smith ML (1994). Measurement-driven Reform: Research on Policy, Practice, Repercussion, CSE Technical Report, 381.Tempe, AZ: Arizona State University, CSE.
- Nunan D (1998). *The Learner-Centered Curriculum*, Cambridge: Cambridge University Press.
- Raza W (2009). English Language Testing in Higher Education of Pakistan, Market Forces. Retrieved from http://www.usm.academia.edu/WajdanRaza/Papers/527642/English_language_testing_in_higher_education_of_Pakistan
- Shamim F (2006). Case-studies of Organization of English Language Teaching in Public-sector Universities in Pakistan, Research Report for the National Committee on English, Higher Education Commission, Islamabad, Pakistan. http://www.aku.edu/sitecollectionimages/aku/zzied/raps/documents/english_in_public_sector.pdf