

Washback and the classroom: the implications for teaching and learning of studies of washback from exams

Mary Spratt *ELT consultant, Cambridge, UK*

This paper reviews the empirical studies of washback from external exams and tests that have been carried out in the field of English language teaching. It aims to do so from the point of view of the teacher so as to provide teachers with a clearer idea of the roles they can play and the decisions they can make concerning washback. The paper begins by defining its use of the term ‘washback’, then goes on to identify the areas in which washback has been noted by the studies. It next examines what intervening factors the studies have indicated influence whether and to what degree washback occurs. This examination highlights how much washback cannot be considered an automatic or direct effect of exams. Finally, the paper pulls together suggestions from the washback literature on how to teach towards exams and indicates areas of classroom practice that these could be applied to. The paper shows how crucial a role the teacher plays in determining types and intensity of washback, and how much teachers can therefore become agents for promoting positive washback.

I Introduction

Madaus (1988: 83) stated that ‘It is testing, not the “official” stated curriculum, that is increasingly determining what is taught, how it is taught, what is learned, and how it is learned.’ This paper reviews recent empirical studies of washback to see whether they indicate this to be the case, and if so, why.

The paper looks at these studies from the point of view of the teacher, whose main concern is generally that of the progress in learning of the group of individuals in their class or classes and

Table 1 Empirical studies on washback

Study	Exam studied	Teaching/learning context	Main issues addressed	Methodology
Alderson and Hamp Lyons, 1996	TOEFL	Language school for university entrants	The influence of TOEFL on classroom teaching	<ul style="list-style-type: none"> • Student interviews • Teacher interviews • Classroom observations
Andrews <i>et al.</i> 2002	Revised Use of English Exam 1994	Hong Kong secondary schools	The influence of the addition of an oral component of the exam on students' oral performance	<ul style="list-style-type: none"> • Video recording of mock oral tests • Grading of oral tests • Discourse analysis
Cheng, 1997; 1998	Revised Hong Kong Certificate of Education Exam in English 1994	Hong Kong secondary schools	The possible washback effect from the revised exam on the teaching of English in Hong Kong secondary schools	<ul style="list-style-type: none"> • Classroom observation of teachers teaching to previous and revised exam • Teacher questionnaires • Student questionnaires
Hamp Lyons, 1998	TOEFL	TOEFL studies	The role of textbooks in test washback	<ul style="list-style-type: none"> • Analysis of 5 TOEFL preparation textbooks
Lam, 1994	Revised Use of English Exam 1989	Hong Kong secondary schools	Whether it is possible in Hong Kong to use the examination system to bring about a positive washback effect on what happens in English language classrooms	<ul style="list-style-type: none"> • Questionnaire to teachers • Analysis of textbooks
Lumley and Stoneman, 2000	Graduating Students' Language Proficiency Assessment	Hong Kong Polytechnic University	Students and teachers' reactions to a learning package of exam preparation materials	<ul style="list-style-type: none"> • Interviews with teachers • Questionnaire to students • Interviews with students and teachers who piloted the package

Read and Hayes, 2003	IELTS	Tertiary institutions and private schools in New Zealand running IELTS or other English classes	The impact of IELTS on preparation for academic study in New Zealand	<ul style="list-style-type: none"> • Questionnaires to schools, teachers and students • Interviews with teachers • Observation of classes • Pre- and post-tests • Student questionnaires • Structured interviews with teachers and inspectors • Analysis of inspectorate bulletins
Shohamy <i>et al.</i> , 1996	English Foreign Language Test Arabic Second Language Test	Secondary schools in Israel	Impact of tests on classroom activities, time allotment, teaching materials, prestige of subject tested, promoting learning Views of different stakeholders: language inspectors, teachers, students	
Turner, 2001	Exams in English as a Second Language	Canadian French-speaking primary and secondary schools	Development of rating scales and its consequential effects on the teachers involved in the development	<ul style="list-style-type: none"> • Feedback from teachers involved in developing rating scales
Wall and Alderson, 1993	Sri-Lankan O-Level Evaluation Project	Sri-Lankan secondary schools	Effects of changing the O-level examinations	<ul style="list-style-type: none"> • Class observation • Teacher and student interviews • Analysis of entrance exam papers • Observation of and interviews with 2 teachers
Watanabe, 1996; 2000	University entrance exams in Japan	Private extra-curricular institutions in Japan preparing students for university entrance exams	Relationship between university entrance examinations and grammar-translation approach to teaching	

their ability as teachers to facilitate that progress. These concerns differ from those of the tester, researcher or educational innovator, whose interests in washback receive attention elsewhere, for example, Bailey, 1999; Wall, 2000.

The term 'washback' is used in the literature with various meanings, which reveal differences in scope, actor and intentionality. The focus of this paper is the classroom. The following definitions of 'washback' capture its meaning as used in the paper:

The influence of the test on the classroom ... this washback effect can be either beneficial or harmful.

(Buck, 1988: 17)

The extent to which the test influences language teachers and learners to do things that they would not necessarily otherwise do.

(Messick, 1996: 243)

The influence of testing on teaching and learning.

(Bailey, 1996: 259)

These definitions focus on the classroom, allow for both the accidental and the intentional effects of washback and leave the door open on whether washback is positive or negative. The term 'washback' will also be used to refer interchangeably to both 'impact' and 'backwash'.

Alderson and Wall in their 1993 article, which put forward various hypotheses on washback, called for empirical research to take place into it: 'Clearly, more research is needed in this area' (1993: 127). The empirical studies since then, which this paper reviews, are given in Table 1.

As will be seen in the following review of the studies, it is still the case that more research is needed on washback, if only to confirm how generalizable the results of these studies are to other populations and situations, and to follow up on issues they raise.

II Areas affected by washback

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. The paper will review the findings for each of these areas in turn.

1 Curriculum

In relation to curriculum, the reports of the effects of washback are contradictory. Alderson and Wall concluded from their Sri Lanka study that 'the examination has had a demonstrable effect on the content of language lessons' (1993: 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 reported an emphasis in teaching on those parts of the exam carrying the most marks. Likewise, Cheng (1997) noted that the content of teaching had changed after the introduction of the revised exam, reading aloud being replaced by role play and discussion activities, for example, reflecting the new exam content. However, Shohamy *et al.*'s 1996 study shows a slightly different picture. They report that the Arabic exam had little effect on the content of teaching whereas the EFL exam did. Watanabe's findings are different again. He speaks of teachers not necessarily teaching listening or writing even though the exam contained these skills. (Watanabe, 1996). The findings of Read and Hayes (2003) are quite detailed and show variations in washback on the curriculum depending on the course observed. Course A was a short, intensive IELTS preparation course; Course B was an extensive one, focusing on general and academic English skills as well as familiarization with IELTS. On Course A, twice as much time was spent on procedural matters as on Course B, 11% of class time was spent on language compared with over one-third in Course B, and almost half the class time was devoted to listening skills, most of which involved listening to the teacher. On Course B, the different language skills were addressed in a more balanced way and greater use was made of integrated skills work.

Other factors relating to the curriculum mentioned in the research are class time allocation and class size. Lam (1994) finds that more curriculum time is given to exam classes, though Shohamy *et al.*'s study suggests 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. Read and Hayes' study (2003) also notes that time allocation may be greater or

lesser depending on the school. They point out too how much of a consideration time is for teachers, with teachers observed remarking that considerations of time available affected their choice of methodology. Alderson and Hamp Lyons also raise the consideration of class size, pointing out that in the situation they investigated there were many more students in the exam classes than in the 'regular' classes.

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 it may not operate at all.

2 *Materials*

The term 'materials' is used here to refer to exam-related textbooks and past papers. Exam-related textbooks can vary in their type of content. 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, emphasizing more the content domain from which the exam is derived. Generally, the studies refer particularly to those materials at the 'highly exam oriented' end of the spectrum.

The 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 coursebooks and other exam materials that are issued when exams are revised. Read and Hayes (2003) confirm this in their New Zealand study, as does Cheng (1997: 50) in Hong Kong: 'By the time the examination syllabus affected teaching in Hong Kong secondary schools ... nearly every school had changed their textbooks for the students'. Shohamy *et al.*'s findings are somewhat different to Cheng's. 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' (1996: 309). However, this is not the case for the low stakes

exam of Arabic for which 'no special courseware ... has been published since 1993' (1996: 304). There seems to be little question that exams generate the publication of exam-related materials if the exams are considered sufficiently high stakes.

How much teachers use these materials seems to vary, however. Lam (1994), though he notes some innovative use of materials generated by the introduction of the revised exam (e.g., the use of teacher-produced authentic materials), also 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 exams is by doing past papers' (1994: 91). 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 note that in 90% of cases in their New Zealand IELTS study, exam preparation books were usually employed (2003).

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.

Watanabe's findings again go against those of others. He found that teachers 'tried to innovate during exam preparation classes ... using a variety of self-made materials' (2000: 44).

The studies indicate that when working towards exams teachers use exam materials to different degrees. Later, we will look at possible reasons for this variation.

A time factor is also mentioned in relation to the use of exam materials: as the exam gets closer there is greater use of past papers and commercial exam-related publications (e.g., Alderson and Wall, 1993).

Two studies contrast teachers' and students' views of exam materials. Lumley and Stoneman (2000) studied teachers' and students' reactions to a learning package for a test newly introduced at tertiary level in Hong Kong and concluded that:

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 saw 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, were 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 they demonstrated relatively little interest in the idea of using test preparation as an opportunity for language learning.

(Lumley and Stoneman, 2000: 75)

This raises the interesting possibility that one reason why some teachers rely heavily on exam-oriented materials may be because they wish to fulfil student expectations or their presumed expectations. Certainly, Alderson and Hamp Lyons (1996: 285) report that the teachers in their study believed it was the students who drove the methodology by insisting on doing practice tests and TOEFL-like items. But the students, who came mainly from Far Eastern countries and Latin America, when asked what they considered the best way of preparing for TOEFL, mentioned 'having American friends', 'going to the movies', 'reading a lot' and generally 'using English outside class'. These students' pedagogic preferences would appear to be very different to those of the Hong Kong students studied by Lumley and Stoneman. A reason for these differences may be the fact that the students in the TOEFL study were actually studying in the United States and therefore

were able to see English in use as a communicative tool, and gain access to it as such to improve their English.

Hamp Lyons looked at the content of exam preparation materials. She carried out a small-scale study of five TOEFL preparation textbooks. She found that 'the skills promoted by the textbooks generally consist of (a) test-taking strategies and (b) mastery of language structures, lexis and discourse semantics that have been observed on previous TOEFLs' and 'the books used for this study promote skills that relate quite exactly to the item types and item content found on the actual test rather than to any EFL/ESL curriculum or syllabus or to any model of language in use' (Hamp Lyons, 1998: 332). This kind of washback on to materials may reflect the kind of exam they focus on, or the teaching and learning context the authors see themselves addressing. Certainly, the author's personal experience of many of the exam preparation materials designed to prepare for various international exams is of a range of kinds of materials, many of which address exams' content domains as much or more than they do exam techniques. Read and Hayes' (2003) findings indicate that the IELTS preparation course made greater use of exam-like materials than the more general course, which used a wider range of texts and materials, including those produced by the students themselves.

What is clear from the above is that both teachers and students behave differently in different learning contexts, and this affects the amount and kind of washback of exams on materials and their use. Studies of the washback from exams onto materials and their users in other teaching and learning contexts would be welcome.

3 *Teaching methods*

By teaching methods I refer to teaching approaches or techniques. The findings on this area are once again not homogeneous. They follow a cline through from indicating no washback to indicating heavy washback.

While Alderson and Wall (1993: 127) say 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) mentioned that teachers make greater use of discussions and role plays after the introduction of the revised exam, but that there is no significant change in the amount of teacher talk. However, although she notes changes in teaching content as a result of the revised exam, she does not observe these changes leading to a change in teaching methods. She comments that the revised exam 'is likely to change the kind of exam practice, but not the fact of the examination practice' (Cheng, 1997: 52), suggesting thereby 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. If an exam changes to become more communicative and the content of teaching changes to reflect this but the teaching methods do not, we have to wonder whether it is the exam itself that is the cause of the change or the lack of change, or whether there are other factors coming into play.

The findings from the Shohamy *et al.* study are once again less clear cut, with the low-stakes Arabic exam involving 'virtually no change from normal teaching' (1996: 304), whereas teaching towards the high-stakes EFL exam led teachers to teach through simulating the exam 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.). The researchers note that these activities become more prevalent the closer the exam gets. These findings are similar to those of Read and Hayes (2003), who found much heavier use of practice tasks, homework and explanation of test-taking strategies in the IELTS preparation course than on the more general course. We should note that while the findings are similar, the variable is not the same in each case. In Israel, it is the nature of the stakes of the exam, in New Zealand the nature of the course. 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: 45).

Further exemplification of the range of ways in which teachers choose to teach towards an exam comes from Smith (1991). In her paper, she reported on a qualitative study of the role of external testing in elementary schools in the USA. As part of the study, she went into schools, interviewed teachers and watched classes in action. She attempted to categorize the kinds of exam preparation that she saw or heard of taking place. Although she watched subjects other than English language being taught, the categories she proposes may prove helpful in facilitating our understanding and awareness of the range of materials and activities used to teach towards exams in EFL exam classrooms. This is how she names and defines her eight categories:

- 1) No special preparation – i.e., no special activities are used to prepare the pupils for the test.
- 2) Teaching test-taking skills – i.e., training testwiseness in skills such as working within time limits or transferring answers to a separate answer sheet.
- 3) Exhortation – i.e., encouraging students to get a good night's sleep and breakfast before the test and to try hard on the test itself. Various forms of prep talks.
- 4) Teaching the content known to be covered by the test – i.e., reviewing the content of ordinary instruction, sequencing topics so that those the test covers would be taught prior to the test, and teaching new content that they know the test covers.
- 5) Teaching to the test – i.e., using materials that mimic the format and cover the same curricular territory as the test.
- 6) Stress inoculation – i.e., test preparation aimed at boosting the confidence of pupils to take the test; working on students' feelings of self-efficacy.
- 7) Practising on items of the test itself or parallel forms.
- 8) Cheating – i.e., providing students with extra time, with hints or rephrasings of words, with the correct answers or altering marks on answer sheets.

(Smith, 1991: 526–37)

Some of the studies indicate that the methods used to teach towards exams vary from teacher to teacher. The studies by both

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. 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. These will be discussed in greater detail below. Besides observing teachers, Alderson and Hamp Lyons interviewed a larger number of them. They conclude from these interviews that teachers did not seem to approach the teaching of their TOEFL classes in the same way as they did non-exam classes. They speak of 'an unteacherly resistance to lesson planning, collecting homework and the other normal professional activities of teaching' (1996: 292). This was in the context of these teachers' general dislike of the 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 but not teaching methods. This perception is not fully supported by these findings. 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 (2000: 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. There is no discussion in the studies of why these findings occur, though we must remember that the TOEFL study indicated that the teachers did not plan their TOEFL classes

as much as they did their others, and also we have seen that exam materials can be heavily used in classrooms particularly as the exam 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. Alderson and Hamp Lyons suggest there may be no uniformity in classrooms in this area either as they reported that the differences between individual teachers teaching towards TOEFL in their use of turn taking, pair work, innovation and laughter are at least as great as the differences between TOEFL and non-TOEFL classes.

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.

4 Feelings and attitudes

Generally speaking, the studies note a gamut of rather negative attitudes and feelings generated by exams. Cheng mentions that students show mixed feelings towards the exam itself, recognizing on the one hand that the exam made them work to achieve good scores but at the same time thinking that exams were not an accurate reflection of all aspects of their study (Cheng, 1998: 296). She also speaks of the pressure felt by teachers, that teachers are worried about how the shy or less outspoken students will fare in the new exam, and of one teacher who admits she would feel guilty if she did not familiarize her students with the test formats. Shohamy *et al.* (1996) find negative feelings towards the Arabic exam and complaints that the test is of no importance. Teachers, however, approve of the EFL exam in as much as they see it as having brought about an acknowledgement of the importance of communicative oral skills that, they believe, will stand their students in good stead in the future. In spite of this, the exam is reported to generate 'an atmosphere of high anxiety and fear of test results among teachers and students. Teachers feel that the

success or failure of their students reflects on them and they speak of pressure to cover the materials for the exam' (Shohamy *et al.*, 1996: 309–10). Similar feelings are also reported by Alderson and Hamp Lyons in the TOEFL study. They say that most of the teachers had a negative attitude towards the exam and to teaching TOEFL, and that they resented the time pressure they felt when teaching towards the exam. Two teachers, however, were much more positive. They 'enjoyed the teaching and felt they could help students cope with something important' (Alderson and Hamp Lyons, 1996: 285). Alderson and Hamp Lyons also mention teachers' feelings of guilt and frustration at 'being unable to make the content interesting or to ensure improved scores for their students' (1996: 292). Finally, they note that there is 'much more laughter in non-TOEFL classes' (1996: 289). This finding is echoed by Read and Hayes who report 1 per cent of class time spent on laughter on the IELTS preparation course and 13 per cent on the more general course. Watanabe also introduces a slightly brighter note. He reports that 'the atmosphere was not necessarily tense. It seemed to depend on the teacher's attitude towards exam coaching' (Watanabe, 2000: 44). Read and Hayes (2003) also report generally positive feelings about IELTS amongst teachers and strong motivation amongst learners.

Once again, it seems that factors beyond the exam itself come into play in determining the amount and kind of washback. In this case they include teachers' attitudes and the stakes of the exam. What these studies do not explore is whether these negative attitudes and feelings generate more or less effective teaching or learning, or whether they impact on them and, if so, how. In the TOEFL study, the quality of the teaching seemed to be negatively impacted by teacher attitudes to the exam, but whether this is the case elsewhere is not clear. Studies of test anxiety and its facilitating or debilitating effects on both teachers and learners during the teaching and learning process merit further research as part of studies of washback. That exams impact on feelings and attitudes seems clear but how these in turn impact on teaching and learning is much less clear.

5 Learning

We come now to those questions about washback that may interest teachers most: does washback from exams affect learning, and, if so, how? Unfortunately, we will see that there is little empirical evidence available to provide a basis for answers to these questions. Wall (2000: 502) said: '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.'

This still seems to be the case three years on. There are some slim findings from English language teaching and other subjects measuring student performance, and within English language teaching, findings on students' learning strategies and performance. Smith, whose study is mentioned above, notes that other researchers found that training in testwiseness had effects that ranged from one-tenth to one-half of a standard deviation on the particular test they investigated, and that test preparation conducted through the use of items of the test itself or parallel forms could inflate scores by six months or more. Of course, we cannot know what these findings might mean for other exams, other learning contexts or other types of exam preparation. Alderson and Hamp Lyons also refer to other researchers' findings on measurement, this time in relation to coaching on SAT scores. They report that 'Powers ... found only dubious evidence for the claims made by coaching companies and test preparation materials publishers that either courses or published materials have any significant effect on students' SAT scores' (Alderson and Hamp Lyons, 1996: 294). The only study amongst those reviewed that attempts to measure learning outcomes is that of Read and Hayes, who make use of retired versions of the IELTS exam as pre- and post-tests. Interestingly, they found that the students on both courses increased their scores and that there was no significant difference in the score increase between the groups. The sample size of their population was, however, small, numbering 17 students in total.

Andrews *et al.* (2002), in their Hong Kong study of the effects of the introduction of a new oral component into a public exam, attempt to measure and describe students' oral performance. They do this by conducting simulated oral tests with three groups of

candidates, matched for their ability. One of these groups had not been prepared for the new test, while the second had taken the test in its first year of implementation and the third in its second year of implementation. The oral tests were then graded by trained examiners. The results indicate a small improvement in performance between the first and the third group. However, the difference is not big enough to be statistically significant, and leads Andrews *et al.* to be very cautious about drawing any conclusions from the results. In the second part of their study the researchers carry out an analysis of the organizational and language features of the candidates' speech by transcribing and analysing their videotaped test data. They report that the third group of students makes greater use of certain discourse markers and times their interviews better than the first group of students. But they also speak of the inappropriate use of these same discourse markers. They conclude:

The sort of washback that is most apparent seems to represent a very superficial level of learning outcome: familiarisation with the exam format, and the rote learning of exam specific strategies and formulaic phrases... the inappropriate use of these phrases by a number of students seems indicative of memorisation rather than meaningful internalisation. In these instances, the students appear to have learnt which language features to use, but not when and how to use them appropriately.

(Andrews *et al.*, 2002: 220–21)

Cheng's Hong Kong study comes up with other negative conclusions: 'The washback effect of this exam seems to be limited in the sense that it does not appear to have a fundamental impact on students' learning. For example, students' perceptions of their motivation to learn English and their learning strategies remain largely unchanged' (1998: 297).

In the Shohamy *et al.* study, teachers report that in their opinion the low stakes Arabic exam may have promoted learning at lower levels but not at upper levels as the students are committed to learning the subject anyway by that stage. In relation to the EFL exam, they believe the new oral component has undoubtedly brought a focus on oral proficiency but that the reading component has not affected reading, as this component of the exam is considered to be poorly designed.

As can be seen, the findings on the washback from exams on to learning are disparate and few.

To summarize this section, we can say that the studies show that there can be washback from exams onto curriculum, materials, teaching methods, feelings and attitudes, and learning. This washback is however not always present and can vary in form and intensity. It seems that other factors beside the exam itself play their part in determining washback. These will be discussed in the next section. We can already see clearly, however, that while the relationship between exams and washback is sometimes thought of as a simple one in which exams generate washback, these studies indicate that rather than there being a direct, automatic and blanket effect of exams, washback is more complex and elusive. It seems to be a phenomenon that does not exist automatically in its own right but is rather one that can be brought into existence through the agency of teachers, students or others involved in the test-taking process.

III The factors that influence washback

Why is it that the occurrence, strength and kind of washback show the variations highlighted above? The factors identified by the empirical studies as influential in affecting washback are many. They can be classified into four main categories: the teacher, resources, the school and the exam itself.

1 Teacher-related factors

In the studies the teacher is constantly mentioned as playing a pivotal role in determining whether washback occurs, how and to what degree. Four main teacher-related factors are cited: their beliefs, their attitudes, their educational level and experience, and their personalities.

In relation to teacher beliefs, the studies mention the following factors as influencing washback: the teacher's beliefs about the reliability and fairness of the exam (Smith, 1991), about what constitute effective teaching methods (Watanabe, 1996), about how much the exam contravenes their current teaching practices (Alderson and Hamp Lyons, 1996), about the stakes and usefulness of the

exam (Shohamy *et al.*, 1996), their teaching philosophy (Lam, 1994), their belief about the relationship between the exam and the textbook (Wall and Alderson, 1993) and their beliefs about their students' beliefs (Alderson and Hamp Lyons, 1996).

Secondly, teachers' attitudes towards the exam: Alderson and Hamp Lyons (1996) mention how teacher attitudes towards the exam affect if and how teachers prepare their classroom materials and their lessons. Turner (2001) reports that teachers' attitudes towards an exam become more positive or promote more positive washback when the teachers are involved in aspects of the test design process.

The third set of factors relates to teachers' education and training. This includes factors such as the teacher's own education and educational experience (Watanabe, 2000), the amount of general methodological training teachers have received (Andrews, 2001), their training in teaching towards specific exams and in how to use exam related textbooks (Wall and Alderson, 1993), their access to and familiarity with exam support materials such as exam specifications, and their understanding of the exam's rationale or philosophy (Cheng, 1997; Wall and Alderson 1993).

Teachers' personalities and their willingness to innovate are also mentioned as intervening factors (Alderson and Hamp Lyons, 1996).

2 *Resources*

The studies mention that resources can affect washback. Factors mentioned are whether or not customized materials and exam support materials, such as exam specifications, are available to teachers (Shohamy *et al.*, 1996; Watanabe, 2000) and the types of textbooks available (Cheng, 1997; Hamp Lyons, 1998).

3 *The school*

Factors mentioned in relation to the school are as follows: its atmosphere and cultural factors such as learning traditions (Watanabe, 2000); how much administrators put pressure on teachers to achieve results (Smith, 1991; Shohamy *et al.*, 1996);

and the amount of time and number of students allocated to exam classes (Alderson and Hamp Lyons, 1996; Read and Hayes, 2003).

4 *The exam*

The studies mention that various factors related to the exam itself can influence degrees and kinds of washback. These include: its proximity, its stakes, the status of the language it tests, its purpose, the formats it employs (Shohamy *et al.*, 1996), the weighting of individual papers (Lam, 1994), when the exam was introduced and how familiar it is to teachers (Andrews *et al.*, 2002).

A summary of the above factors in the form of a list has been included in Appendix A.

The factors focus on the individual teacher and on the teacher as part of a wider system. Teachers, like everyone else, operate in ideological, historical, economic and political contexts that affect their attitudes, beliefs and behaviours.

The studies to date do not show in what directions the factors push washback. For example, would a well-trained and educated teacher working with an exam of which he or she approved and about which he or she was well informed be more or less likely to adhere to the content of the exam in their lessons? The studies indicate that the answer to this question would likely be: it depends. There is also an interaction between the factors and between the factors and the teaching and learning contexts, which is not as yet described. The variety of the factors, their varying strength and the complexity of the interactions between them indicate strongly that washback does not always occur and that when it does it may do so in a variety of forms and intensities in different contexts.

IV Guidelines on teaching towards exams

It can be concluded from the studies that washback is not inevitable and also that it is malleable. This conclusion puts the teacher in the driving seat in some important ways as far as washback is concerned. When and where the teacher is in control of the factors determining washback, washback itself is largely in the teacher's

control. It is the teacher who can then determine to a greater or lesser extent whether to allow washback to operate, what areas it should operate in and how. This means the teacher has a series of decisions to make, decisions both pedagogic and ethical. Possible parameters for these decisions and the areas they operate on will be discussed below. They are based on suggestions from the literature on washback, and the pointers emerging from the above review. They aim to facilitate positive washback.

The decisions a teacher needs to make concerning teaching towards exams involve choices about the best ways of teaching and promoting learning to achieve both good exam results and good learning of the content domain of a syllabus. With some exams or administrative arrangements for courses, however, a teacher may note a conflict between teaching and learning requirements and exam success requirements. This conflict can create a tension between pedagogical and ethical decisions. Bailey (1996) and Hamp Lyons (1998), amongst others, point out that the tension between pedagogic and ethical decisions occurs when teachers believe that 'tests run contrary to the principles and practices of current approaches to language learning' (Bailey, 1996: 259), and when they believe that the most effective way for their students to achieve higher test scores is to be given opportunities to engage in some form of test coaching. We should remind ourselves at this point that, as pointed out above, there is currently no evidence that test coaching achieves better test scores.

Hamp Lyons argues for a code of ethical practice for those involved in test preparation. She suggests that practice on published previous or parallel forms is both educationally indefensible as it boosts test scores without mastery, and of dubious legality as it coaches merely for score gain (Hamp Lyons, 1998: 334). This position echoes that of various educationalists. Smith reports that Cannell, for example, imputes that any test preparation practice that artificially inflates scores and thereby robs the public of accurate information is immoral (Smith). Smith herself, however, points out that many teachers view the use of practice materials and activities differently from some educationalists, as they do not believe in the inherent reliability of a test as a true reflector of

student performance. This is because of the possible presence of various kinds of bias in a test, for example, bias against students from particular socioeconomic or ethnic backgrounds or against those with a particular emotional make-up.

The tension between ethical and pedagogic decisions reflects itself in the following suggestions. McKay (2001) suggests a practical method for resolving tension between the demands of short and long-term assessment. This could be adapted in certain learning contexts to resolve tensions between ethical and pedagogic decisions related to assessment. She argues for adapting the idea promoted by Woodward (1988) of 'pushing' and 'popping'.

'Pushing' in computer programming language is suspending operations on the task currently being engaged in and taking up a new task. This task is usually said to be on a lower level than the first task. Once this second lower level task is completed the teacher can 'pop' back up to the first level again.

(McKay, 2001: 24–25)

We can think of the top-level task as helping students to learn a content domain, and the lower level one as helping students to pass an exam successfully.

Bailey suggests four ways of reducing tension between pedagogic and ethical decisions and of promoting beneficial washback. These are 'the incorporation of 1) language learning goals; 2) authenticity; 3) learner autonomy and self-assessment; and 4) detailed score reporting' (1996: 268).

Bailey's suggestions appear to be written for an audience of test writers but they can also be adapted by the teacher to guide choices in the classroom. With regard to educational goals, she says that 'Washback can either be positive or negative to the extent that it either promotes or impedes the accomplishment of educational goals held by learners and/or programme personnel' (Bailey, 1996: 269). The teacher can incorporate this suggestion by ensuring that educational goals are pursued in the classroom. In relation to authenticity, she refers to the use of both authentic tasks and authentic texts in testing. While teachers may not be able to control the content of external exams, they could apply this advice in their classrooms in that they often control the content of the class-based tests they employ to teach towards an

external exam as well as the texts and tasks they use for teaching towards the exam's content domain. In relation to learner autonomy and self-assessment, Bailey suggests enabling students to assess their own abilities and being given responsibility for doing so. This advice can be adapted for the classroom. Finally, with regard to score reporting Bailey suggests that exam boards provide full feedback on test performance. The teacher could adapt this advice to ensure that they themselves provide full feedback to students on class tests.

The above suggestions from McKay and Bailey provide general guidelines for principled approaches to decisions about teaching towards exams. They can complement principles coming from theories of language teaching and learning. The empirical studies reviewed in this paper identify specific points within the areas of teaching and learning that are susceptible to washback. The teacher could apply the guidelines to these points. To conclude this article the points are summarized below under the areas discussed:

- *Curriculum* – how much to focus on the exam's content domain as opposed to exam techniques and test wiseness, when to teach particular areas of the curriculum, how much time to devote to teaching particular areas.
- *Materials* – what textbooks to use, how much use to make of selected textbooks, how much and how to use exam or parallel exam materials, how much to use other materials including one's own and the students'.
- *Teaching methods* – how much drilling to employ, when to employ such methods, how much to employ other methods more focused on language development and creativity, what kinds of exam preparation to employ (cf. Smith's eight categories), how much planning time to devote to exam classes, what kind of atmosphere to promote in exam classrooms, what kind of interaction patterns to encourage in exam classrooms.
- *Feelings and attitudes* – what kinds of feelings and attitudes towards the exam to attempt to maintain and promote in students.
- *Learning* – the appropriateness of the learning outcomes demonstrated by students.

V Conclusions

This review of recent empirical studies of washback shows that the number of studies remains relatively small, that they have been carried out in a restricted number of learning contexts and have employed a variety of research methods. There is a need for more studies to be carried out in different learning contexts. Use of parallel methodologies for studies in different contexts might also allow researchers to investigate some of the apparent contradictions in the findings to date.

Nevertheless, the empirical studies reviewed indicate strongly that an exam cannot of itself dictate what and how teachers teach and learners learn. Degrees and kinds of washback occur through the agency of various intervening bodies and are shaped by them. An important and influential agent in this process is the teacher. This suggests that teachers face a set of pedagogic and ethical decisions about what and how best to teach and facilitate learning if they wish to make the most of teaching towards exams. In saying this, it should not be forgotten that the teacher in the classroom operates within an ideological, historical, economic and political context.

VII References

- Alderson, C.** and **Hamp Lyons, L.** 1996: TOEFL preparation courses: a study of washback. *Language Testing* 13(3): 280–97.
- Alderson, C.** and **Wall, D.** 1993: Does washback exist? *Applied Linguistics* 14(2): 115–29.
- Andrews, S.** 1995: Washback or washout? The relationship between exam reform and curriculum innovation. In Nunan, D., Berry, R. and Berry, V., editors, *Bringing about change in language education*. Hong Kong: Department of Curriculum Studies, University of Hong Kong.
- 2001: Reflecting on washback: high stakes tests and curriculum innovation. Paper given at ILEC Conference, Hong Kong.
- Andrews, S., Fullilove, J.** and **Wong, Y.** 2002: Targetting washback – a case study. *System* 30(2): 207–23.
- Bailey, K.** 1996: Working for washback: a review of the washback concept in language testing. *Language Testing* 13(3): 257–79.
- 1999: Washback in language testing. *TOEFL Monograph Series* MS-15. Princeton, NJ: Educational Testing Service.

- Buck, G.** 1988: Testing listening comprehension in Japanese university entrance exams. *JALT Journal* 10: 15–42.
- Cheng, L.** 1997: How does washback influence teaching? Implications for Hong Kong. *Language and Education* 11(1): 38–54.
- 1998: Impact of a public English examination change on students' perceptions and attitudes toward their English learning. *Studies in Educational Evaluation* 24(3): 279–300.
- Hamp Lyons, L.** 1997: Washback, impact and validity: ethical concerns. *Language Testing* 14(3): 295–303.
- 1998: Ethical test preparation practice: the case of the TOEFL. *TESOL Quarterly* 32(2): 329–37.
- Lam, H.P.** 1994: Methodology washback – an insider's view. *Bringing about change in language education*. Hong Kong: Department of Curriculum Studies, University of Hong Kong, 83–99.
- Lumley, T. and 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.
- McKay, P.** 2001: Innovation in English language assessment: looking towards long-term learning. In Davison, C., Crew, V. and Hung, J., editors, *Conference Proceedings. International Language in Education Conference, 2000*, Hong Kong: University of Hong Kong [CD ROM].
- Madaus, G.F.** 1988: The influence of testing on the curriculum. In Tanner, L.N., editor, *Critical issues in curriculum: eighty-seventh yearbook of the National Society for the Study of Education* (Part 1), Chicago: University of Chicago Press.
- Messick, S.** 1996: Validity and washback in language testing. *Language Testing* 13(3): 241–56.
- 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.
- Shohamy, E., Donitsa-Schmidt, S. and Ferman, I.** 1996: Test impact revisited: washback effect over time. *Language Testing* 13(3): 298–317.
- Smith, M.L.** 1991: Meanings of test preparation. *American Educational Research Journal*, 28(3): 521–42.
- Turner, C.** 2001: The need for impact studies of L2 performance testing and rating: identifying areas of potential consequences at all levels of the testing cycle. In Elder, C., Brown, A., Iwashita, N., Grove, E., Hill, K. and Lumley, T. editors, *Experimenting with uncertainty: essays in honour of Alan Davies*, Cambridge: Cambridge University Press, University of Cambridge Local Examinations Syndicate.
- Wall, D.** 2000: The impact of high stakes testing on teaching and learning – can this be predicted or controlled? *System* 28(4): 499–509.

- Wall, D. and Alderson, J.C.** 1993: *Examining washback: the Sri Lanka impact study*. *Language Testing* 10(1): 41–69.
- Watanabe, Y.** 1996: Does grammar translation come from the entrance examination? Preliminary findings from classroom-based research. *Language Testing* 13(3): 318–33.
- 2000: Washback effects of the English section of Japanese entrance examinations on instruction in pre-college level EFL. *Language Testing Update*. 27 (Summer): 42–47.
- Woodward, T.** 1998. *Loop input*. Exeter: Pilgrims' Publications.

Appendix A: Factors identified by empirical studies as affecting degrees and kinds of washback

Teacher-related factors

Teacher beliefs about:

- the reliability and fairness of the exam
- what constitute effective teaching methods
- how much the exam contravenes their current teaching practices
- the stakes and usefulness of the exam
- their teaching philosophy
- about the relationship between the exam and the textbook
- their students' beliefs

Teachers' attitudes towards:

- the exam
- preparation of materials for exam classes
- lesson preparation for exam classes

Teachers' education and training:

- Teachers' own education and educational experience
- the amount of general methodological training they have received
- training in teaching towards specific exams and in how to use exam-related textbooks
- access to and familiarity with exam support materials such as exam specifications
- understanding of the exam's rationale or philosophy.

Other:

- personality
 - willingness to innovate
-

Resource, the school, the exam

Resources:

- the availability of customised materials and exam support materials such as exam specifications
- the types of textbooks available

The school:

- its atmosphere
- how much the administrators put pressure on teachers to achieve results
- the amount of time and number of students allocated to exam classes
- cultural factors such as learning traditions

The exam:

- its proximity
- its stakes
- the status of the language it tests
- its purpose
- the formats it employs
- the weighting of individual papers
- when the exam was introduced
- how familiar the exam is to teachers