Glover, P. (2014). Do language examinations influence how teachers teach? *International Online Journal of Education and Teaching (IOJET), 1*(3). 197-214. http://iojet.org/index.php/IOJET/article/view/48

DO LANGUAGE EXAMINATIONS INFLUENCE HOW TEACHERS TEACH?

Philip Glover

Süleyman Demirel University

pgaglover@gmail.com

Biodata

Assistant Professor Dr. Philip Glover teaches in the English Language and Literature Department at Süleyman Demirel University in Isparta. He has taught English and trained English teachers in Turkey, Cyprus, UK and Hungary. He has worked in English language teaching as a manager and as a consultant with testing and in-service training projects. His research interests include methodology, classroom discourse, language testing and intercultural learning.

Copyright by Informascope. Material published and so copyrighted may not be published elsewhere without the written permission of IOJET.

DO LANGUAGE EXAMINATIONS INFLUENCE HOW TEACHERS TEACH?

Philip GLOVER
pgaglover@gmail.com

Abstract

Examination influence on teaching, known as washback, has been found on curriculum, materials and attitudes to teaching. Evidence of washback on how and why teachers teach, however, has proved harder to identify. This paper looks in detail at three complicating factors in washback research; the variety of terms used, conflicting findings and the nature of positive or negative washback. The paper describes a study that used discourse analysis in order to investigate examination influences on teacher talk. Findings suggest that washback on how teachers teach, if present, may lie in some categories of teacher talk but not in others.

Keywords: language testing, examination washback, language teaching

1. Introduction

Washback, 'the effect of testing on teaching and learning' (Hughes, 1989, p.1), is change in teaching caused by an examination, and is of interest to teachers, administrators and innovators. Some teachers and testers believe that tests can exert a powerful influence on teaching, and may be harnessed to raise standards of teaching and learning. Thus 'measurement-driven instruction' (Popham, 1987) has the hope and expectation that testing will 'shape and pull teachers' practices in desirable ways' and 'motivate teachers to improve their teaching' (Chapman & Snyder, 2000). Others have observed that examinations have negative effects on the curriculum taught (Madaus, 1988), on problem-solving skills (Frederiksen, 1984) or on time spent teaching (Smith *et al*, 1991). There are examples from around the world of success and failure in the use of tests to raise standards of teaching (Wall, 2000; Kellaghan & Greaney, 1992).

In the field of language testing, assumptions about washback were challenged by Alderson and Wall (1993, p.115), who observed that: 'very little evidence has been presented to support the argument that tests influence teaching'. They proposed washback hypotheses for investigation, including hypotheses that tests may influence what teachers teach, how teachers teach and attitudes.

Subsequent research in different parts of the world found evidence of washback on what teachers teach and attitudes in the form of:

- More attention to parts of the curriculum tested in the examination, with activities in class showing wider, narrower or simply different teaching content in examination lessons;
- Materials used for teaching that reflect the content of the examination;
- Teachers' often negative attitudes to what and how they teach for examinations.

A number of studies identified washback on curriculum content and classroom activities. Wall and Alderson (1993) and Wall (1999) found curriculum-narrowing with the content of

teaching limited to what was tested, more time being spent in lessons on writing and reading. Alderson and Hamp-Lyons (1996) found teachers reported effects such as curriculum-narrowing, lost instructional time and reduced emphasis on skills that require complex thinking. A study of 'assessment-driven reform' found that the inclusion of a writing test increased attention to writing (Stecher *et al.*, 2004, pp.68-69). Washback on classroom activities was also found in Japan (Watanabe, 1996) and in New Zealand (Hayes & Read, 2004).

Examination influence on materials was observed by Wall (1999), who found teachers used supplementary books in examination preparation to compensate for lack of grammar in the course book. Shohamy *et al* (1996) found teachers of a new Arabic-as-a-second-language examination replaced textbooks with exam-type sheets. Cheng (1997) found that teachers relied on textbooks to interpret the new examination and also found that materials changed as new books were introduced. Differences in language produced by students in tests were attributed to the influence of materials published for a new examination (Andrews *et al.*, 2002), and Nikolov (1999, p.243) found washback on supplementary materials. Watanabe (1996) found textbook materials used in class were past papers or constructed by teachers on the examination model. In a study of Cambridge Proficiency Examination study evaluators found 'books tend to represent directly the content, approaches, activities and tasks of the exam' (Hawkey, 2004).

Evidence of washback on attitudes has also been found, often as a conflict between how teachers would like to teach and how they feel they are forced to teach for examinations (Smith *et al,* 1991, p.41). Hughes (1989, p.1) refers to a writing skills course tested by multiple choice which leads to 'pressure to practise such items rather than practise the skill of writing itself'. Prodromou (1995, p.14) claims that 'sound teaching practices are often sacrificed in an anxious attempt to 'cover' the examination syllabus, and to keep ahead of the competition'. Alderson and Hamp-Lyons (1996, p.285) found negative attitudes to examination teaching, with teachers complaining that teaching became 'boring and fragmentary', resenting time pressure. Shohamy *et al* (1996, pp.308-9) found teachers claimed that an English examination forced them to teach in a certain way or felt they could teach more creatively at times when not under examination pressure whilst others appreciated the motivation provided by the oral test. Qi (2005, p.154), suggests that a test's 'selecting and evaluative functions lead to the short-term goal of teaching to raise scores', which works against teachers' long-term goal of improving language proficiency in class. Others refer to negative attitudes to what and how teachers have to teach for examinations (Kiss-Gulyas, 2001; Cheng, 1997). There is an overlap in research between studies of how teachers teach and attitudes to teaching, with self-reports about teaching behaviour being taken as evidence of classroom practice (Stecher *et al,* 2004), rather than as evidence of attitudes.

1.1. Washback on How Teachers Teach

Research into washback on how teachers teach is more complex than that for what teachers teach and for attitudes. There seems to be a conflict between on the one hand claims that 'most teachers are familiar with the amount of influence testing can have on their instruction' (Bachman and Palmer, 1996, p.33) or that 'there is a general consensus that high-stakes tests produce strong washback' (Qi, 2005, p.3), and on the other hand empirical studies which conclude that 'there is no evidence of 'washback' on methodology' (Wall & Alderson, 1993, p.66), or that after the introduction of a new test; 'it can be seen that the general pattern of teaching approaches had not changed much' (Cheng, 1999, p.268), or 'the use of achievement tests has no clear influence on teaching practices' (Wesdorp, 1982, p.48).

Three factors complicate interpretation of washback research into how teachers teach:

Factor 1: the variety of terms used and classroom features investigated.

Factor 2: conflicting findings.

Factor 3: positive or negative examination effects do not relate to theories of good teaching.

These ontological and epistemological issues may have contributed to some of the ‘apparent contradictions’ (Spratt, 2005, p.27) that have been perceived in washback studies. Diversity of terminology and a focus on varying aspects of classroom teaching may create difficulties in comparing studies. Some studies found no washback effects on teaching, others found washback to be present and still others found washback was varied, present in some ways but not in others. Table 1 (adapted from Author, 2006) summarises 16 washback studies in terms of findings, use of observation, terms and classroom features.

Table 1. *Studies, Terms, Aspects Investigated, Observation and Washback*

Study	Terms used	Classroom Aspects investigated	Observation	Washback
Wall (1999)	Methods, methodology	Explaining words and structures	Yes	Absent
Wall and Alderson (1993)	Methodology	Pre-teaching vocabulary for reading, Reading skills work	Yes	Absent
Wesdorp (1982)	types of teaching, teaching practices	Explaining, Demonstrating, Interaction patterns.	No	Absent
Alderson and Hamp-Lyons (1996)	Methodology, teaching practices	Talking time, Test-taking time, References to the examination, Laughter, Innovations, Metalanguage, Pair work	Yes	Variable
Watanabe (1996)	Methods	Translation, Grammatical explanations	Yes	Variable
Cheng (1999)	Tasks	Integrated language tasks, Explaining mock exams, Group work	Yes	Variable
Hayes and Read (2004)	Activities	Test-taking activities, Interaction, Feedback, Explanations, Student strategies, Laughter	Yes	Variable
Burrows (2004)	Methods	Curriculum, Teacher discourse, Explanations, Instructions, Interaction.	Yes	Variable
Nikolov (1999)	Activities, techniques, tasks	Translation, Gap-filling, Reading aloud, Grammar exercises	Yes	Variable
Cheng (1998)	Activities	Reading aloud, role play, group discussion	No	Variable
Andrews (1995)	Pedagogical strategies	Speaking for presentations, Group discussions, Using set phrases, Grammatical accuracy	No	Variable
Shohamy et al	Methodology,	Test activities, Memorization,	No	Variable

(1996)	content, Activities	Speaking at length, Group and pair work, Debates, discussions		
Chapman and Snyder (2000)	Cognitive load, Strategy	Multiple choice/ short answer questions, Problem-solving, Critical thinking, Rote learning	No	Variable
Stecher et al (2004)	Classroom practice, Methods	Explaining, Suggesting revisions, Giving examples, Discussion.	No	Present
Turner (2001)	Methodology, tasks	Performance tasks, Teacher discourse	No	Present
Prodromou (1995)	Methods	Penalising error, questions, Denying communication, Anxiety, Solemnity.	No	Present

1.1.1 Complicating Factor 1: The Variety of Terms and Features Studied

The 16 papers in Table 1 use 12 different terms for classroom teaching. It is not clear whether 'Method', 'methodology', 'practice', 'technique' and so on reflect fundamental differences in the authors' views of classroom events, or whether they may simply reflect a diversity of terminology used in English language teaching to describe similar phenomena. The authors in Table 1 do not state that they use their chosen term in order to differentiate their study from each other, but choose the word to refer to specific features of classroom teaching that they investigated. The terms are not defined, nor are differences between 'method', 'methodology', 'practice' or 'technique' explored. They do not imply distinctions such as between approach, method and technique (Anthony, 1963) or between approach, design and procedure (Richards & Rodgers, 1986, p.28), The terms relate to what happens in the classroom, the technique or procedure level of techniques, practices and behaviours that are observable when a method is employed. The choice of term therefore is not central to the studies, and the terminology does not reflect fundamental conflicts of principles, but rather linguistic choices available to describe similar events.

Whilst the use of so many terms does not seem to prevent comparability, the use of the term 'method', (Wall, 1999; Watanabe, 1996; Burrows, 2004; Stecher *et al,* 2004; Prodromou, 1995) may be problematic. In English language teaching, what actually constitutes a 'method' is not clear and is open to interpretation. 'Method' as a discrete set of procedures for a specific teaching purpose has been rejected as unhelpful (Prabhu, 1990), or even pronounced dead (Allwright, 1991). According to these views 'method' is a personal construct describing how an individual teacher teaches rather than a set of procedures that a teacher may choose to employ in lessons. If 'method' is a personal construct, then the use of the term itself may lead classroom research to highlight the individuality of teaching rather than common features in how teachers teach. Research comparing different methods in English language teaching encountered a similar difficulty (Oskarsson, 1973), where attempts to compare different 'methods' were confounded by teachers' individual differences. The study reported in this paper refers instead to 'how teachers teach' in order to avoid the problematic nature of the term 'method', and in order to bring together the different terms for teaching into one broad category.

There are 58 aspects of classroom teaching investigated in the studies, but they have much in common and relate to different aspects of discourse and classroom practices. Studies of discourse have looked at teacher and student talk, the number of turns, references to the test, metalanguage use (Alderson & Hamp-Lyons, 1996), explanations, discourse and interaction

(Burrows, 2004), grammatical explanation (Watanabe, 1996), talking about aims, explaining mock exams, textbook exercises, language items, meaning (Cheng, 1998, 1999), the nature of the feedback given, the use of explanations (Hayes & Read, 2004) explanations, instructions, discourse (Burrows, 2004), student talk (Andrews, 1995, p.79; Shohamy *et al,* 1996, p.301) and explaining vocabulary (Wall, 1999). All these studies view aspects of classroom talk as possible evidence for washback.

Studies of practices have looked at pair work, innovations (Alderson & Hamp-Lyons, 1996), translation (Watanabe, 1996), integrated language tasks, group work (Cheng, 1998, 1999), focus on student strategies (Hayes & Read, 2004), memorization, speaking at length, group and pair work, debates and discussions. (Shohamy *et al,* 1996, p.301), reading skills work and pre-teaching vocabulary (Wall & Alderson, 1993). All these studies see how activities are carried out (individually, in pairs or groups, with the teacher, in a certain atmosphere) as possible evidence for washback.

1.1.2. Complicating Factor 2: Conflicting Findings

Some studies found washback to be absent, some found washback varied between individual teachers or between different features of classroom teaching and some found washback to be present. These conflicting findings may be explained by looking at the nature of the evidence considered and the role of factors such as context. One possible explanation of conflicting findings lies in what is taken as evidence of how teachers teach. Studies that claim to have found evidence of washback have tended to use report data from teachers instead of observation. Studies that elicited reports on classroom behaviour without observing classrooms may indicate washback on attitudes rather than washback on how teachers teach.

Another source of conflicting perceptions is between those studies that found no washback and those that found evidence varied. A major study in Sri Lanka (reported in Wall and Alderson, 1993; Wall, 1999; Wall, 2005) concluded that washback on how teachers teach was not present. This study made extensive use of classroom observation and found that, whilst the content of teaching, materials and attitudes were influenced, the teachers taught for the new examination in the same way as they had done for the old one. Before and after the innovation the teacher's role was to 'make the child understand' rather than promote learning (Wall, 2007, p.147). The study identifies teacher cognition and resources as key factors in the innovation: 'the exam can have no impact on methodology unless the teachers understand correctly what it is the exam is testing' (Wall & Alderson, 1993, p.65), which is attributed to lack of resources for training teachers. Other studies found washback varied between one teacher and another (Watanabe, 1996) or varied in extent between teachers, more for one and less for another (Alderson & Hamp-Lyons, 1996). Other studies found washback of limited extent (Cheng, 1999) or varying according to context (Chapman & Snyder, 2000). According to these studies examination influence on how teachers teach was present but variable. Factors such as teacher cognition or time since the innovation were more influential than the examination itself. As different findings have emerged in different contexts, there seems to be no universal examination effect on the classroom, and findings should not be dismissed as contradictory, but rather as context specific. Washback may be found in specific contexts, or in different teachers, and washback is therefore worth researching.

1.1.3. Complicating Factor 3: Positive or Negative Effects

A third problem in washback studies is the identification of positive or negative influences on teaching. Messick (1996) cautions that negative washback should not be confused with ‘bad teaching’, noting that claims of negative washback need to be supported by evidence that examination lessons are indeed different, that any difference is the result of the

examination, and that any difference represents a negative effect on the teaching. Aspects of how teachers teach that are investigated need to be clearly associated with positive or negative influences on learning.

Some studies have referred to practices such as pair or group work or discourse patterns such as 'IRF'. Whilst the presence of pair work is likely to make a positive contribution to learning, more pair work is not necessarily better than less (Edwards & Mercer, 1987; Hay McBer, 2000). Consequently the presence or absence of pair and group work may show positive or negative washback, but more or less pair and group work does not increase the effect. The idea that student-student interaction makes a greater contribution to learning than teacher-student interaction is not supported by research that indicates the important contribution of teacher-led interaction (Muijs & Reynolds, 2005). Similarly positive or negative effects may be found in the presence or absence of classroom practices such as translation, integrated language tasks, focus on student strategies, memorization, speaking at length, debates, discussions, frequency of shared laughter, reading skills work or pre-teaching vocabulary, but it is difficult to claim that more or less of these features increases any positive or negative effects.

As for discourse, several washback studies refer to the 'IRF' pattern and its assumed negative consequences for teaching and learning (Cheng, 1999; Nikolov, 1999; Shohamy *et al.*, 1996), using 'IRF' as an indication of conventional, teacher-centred teaching. 'IRF' is associated with the old examination for Nikolov, and Cheng expresses disappointment that a new examination has not broken the 'IRF' pattern. Classroom interaction has been viewed in terms of its contribution to learning by facilitating lifelike or 'genuine' communication (van Lier, 1988, p.28), and the presence of 'IRF' patterns in the classroom (Sinclair, 1982) has been used to analyse and comment on many aspects of language teaching (Cadorath & Harris, 1998; Cullen, 1998; Dinsmore, 1985; Duff, 2002; Hall, 1998; Kumaravadevilu, 1993; Nunan, 1987; Ohta, 1999; Thornbury, 1996). The presence of 'IRF' patterns in lessons has been used to support a view that some classroom interaction does not replicate real-life communication and as a consequence does not facilitate language learning. Alternative views of classroom discourse, however, see talk in lessons in a different light, with the function of guiding and supporting learning. According to this view it is more important to understand how 'IRF' patterns 'relate to the core institutional goal rather than dismissing them as undesirable or not genuine' (Seedhouse, 1996, p.22), and McDonough (2002, pp.138-139), suggests there is a need to 'de-couple' natural language learning processes from classroom interaction in order to see the classroom as another 'natural' environment which has its own language with its own purposes and characteristics.

1.2 Conclusions and Expectations about Washback on How Teachers Teach

There seem to be sufficient common features in the studies above to draw strong conclusions about washback as it has been investigated so far:

- Washback can influence how teachers teach but effects are limited by other influences such as context, resources and teacher cognition, leading to variations between teachers and situations.
- Examinations influence how teachers teach in some situations and for some teachers.
- The nature of such an influence needs further investigation in order to build on findings from previous research.
- Previous research has investigated effects on classroom discourse and practices, but a detailed description is lacking of what washback may look like in the classroom.

1.3. Expectations for Washback Based on Previous Research

1. Differences are likely to be found in examination lessons in what teachers teach and attitudes. Differences in aspects of how teachers teach, if present at all, are likely to vary between teachers and contexts.
2. Washback on how teachers teach might be present in classroom discourse or practices, and investigated through the use of a combination of observation, discourse analysis and teacher reports.
3. Positive or negative effects may be indicated by differences in classroom events and their positive or negative contribution to teaching.

Teachers have been exhorted to 'work for washback' (Bailey, 1996) and administrators expect changing an examination to change teaching. It is suggested that teachers can determine the extent of examination influence (Spratt, 2005). A clearer idea of what washback looks like may help teachers, administrators and students to work for the desired improvements. Findings from the study reported below may indicate aspects of how teachers teach that may be affected by an examination.

2. The study

The study (Author, 2006) investigated washback by looking in detail at classroom discourse in the form of teacher talk. The study focused on analysis of classroom discourse supported by field notes from lesson observation and teacher reports elicited by interview and questionnaire. The study made a detailed analysis of teacher talk using an analytical framework, firstly to identify aspects of teacher talk that were different in examination lessons, secondly to consider what differences show about how teachers teach, and thirdly to see whether these differences represent positive or negative washback. The study asked:

1. Does the teaching show evidence of possible washback?
2. What does possible washback show about how teachers teach in examination lessons?
3. Do differences in how teachers teach show positive or negative effects?

Classroom discourse data for the study came from transcriptions of teacher talk recorded in lessons, field notes and teacher reports from 12 English language lessons taught by two teachers teaching four groups in Central Europe over a period of 10 days. Eight hours of recordings, made with a lapel microphone carried by the teacher, were transcribed, producing transcriptions of 23,506 words used by the teachers. Comments on the teacher talk were added in brackets describing non-verbal sounds such as laughter and writing on the board, but prosodic features, pauses and so on were not included, and the transcriptions can be described as 'verbatim' (Rose, 2000, p.250).

In half of the twelve lessons final (twelfth) year students were preparing for the English language section of the school-leaving examination, and in the other six lessons students were in the eleventh year following a general English program. The study maximised comparability of the lessons by observing teachers and classes that were as similar as possible; two teachers with the same subject specializations, similar qualifications and experience in the same school teaching examination and non-examination classes in the same year with classes that were parallel in terms of age and academic program. These steps aimed to reduce the possibility that differences between the lessons may be due to factors unrelated to the examination.

The examination certifies the completion of high school studies, Years 9 to 12. Passing is a requirement for certain jobs and for entry into tertiary education, so the examination is a

high-stakes examination because of its serious consequences for the test takers, with failure preventing access to higher education and some jobs. A very high pass rate in the subject examinations somewhat detracts from this high-stakes status (Ábrahám & Jilly, 1999, pp.35-36). Nevertheless, consequences for the students who failed were serious, and the examination was a challenge for many candidates and important for school authorities (Alderson & Szollás, 2000, pp.17-21). The test takers were school leavers who had opted for the English language examination but who had not gained an exemption by passing the intermediate level state language examination. Candidates therefore tended to be the less successful learners of English (Ábrahám & Jilly, 1999, p.21). The washback of the examination was thought by teachers to be widespread and negative (Kiss-Gulyás, 2001, p.43). A study of 118 lessons found evidence of washback on materials, lesson content and tasks (Nikolov, 1999).

2.1. Supporting Data: Field Notes and Reports

The supporting data produced evidence that the examination lessons were linked to the examination and showed the presence of washback on what teachers teach and attitudes. The curriculum in the examination lessons was clearly based on the examination. Both teachers covered examination content in the examination lessons and language forms and skills in the non-examination lessons. Preparation for the examination consisted of practising examination tasks in lessons. Examination tasks were an oral presentation and interview, translation and a reading and grammar test. These activities were observed in lessons and reported by teachers who expressed their lesson aims in terms of examination tasks in examination lessons and language forms (grammar, functions or vocabulary) in non-examination lessons. Materials used in the examination lessons were taken mainly from past papers, whereas a general English course book was used in non-examination lessons.

Teachers' attitudes were also affected by the examination. Both teachers expressed negative views about the format of the examination and the fact that they felt obliged to abandon regular teaching in order to prepare students for the examination. They also commented on negative effects on student motivation caused by the pressure of the examination on the students.

The use of pair and group work, however, did not seem to be affected by the examination. Both teachers used pair and group work in both types of lesson, even though there was no pair or group interaction in the examination. These teacher choices of interaction patterns seem to have been motivated by a belief that such activities made a contribution to learning in general.

2.2. Teacher Talk Data

The study analysed discourse in examination lessons by looking at teacher talk. Classroom discourse can provide insights into how teachers teach in examination lessons because discourse is action through talk that is analyzable, is variable between people and situations and is interpretable as showing purpose. Discourse requires detailed analysis and can 'produce social explanations which are generalisable in some way, or which have a wider resonance' (Mason, 1996, pp.4-6), for example the consequences of examinations. Even a small-scale study of classroom discourse may show action that is repeated by many, though not necessarily all or most teachers.

In English language teaching discourse analysis has produced insights into classroom talk and how teachers teach, for example Duff (2002) looked at issues of identity, respect and language socialisation in a school in Canada, focusing on two lessons and producing insights into relations between students and the teacher's attempts to make cultural connections. Another study investigated student talk whilst preparing for and carrying out a task relate to

language learning (Mori, 2002) and yet another used systemic functional grammar to provide insights into science teaching (Young & Nguyen, 2002) by comparing language used by a teacher in a lesson with language in a text book.

A specific characteristic of language used by teachers is its pedagogical purpose, to facilitate learning. Mercer (1995, p.1) describes classroom communication 'in which one person helps another to develop their knowledge and understanding' as 'the guided construction of knowledge'. The role of the teacher can be 'crucial' to the success of learning, and Mercer notes classroom research identifying ways for teachers to guide learners more effectively (Mercer, 2000, pp.159-165). This guiding is described as 'scaffolding', which is seen as 'a useful metaphor for the intellectual involvement of a teacher with a learner's efforts during joint activity' (Mercer, 2000, pp.169-170). A range of different viewpoints have stressed the contribution of teaching to learning. The Russian psychologist Vygotsky, writing in the thirties saw the support and guidance that teachers provide in lessons through language as operating in a 'zone of proximal development' to facilitate learning (Vygotsky, 1978, pp.84-86). Others have cited a range of language factors that contribute to learning such as explaining, questioning (Muijs & Reynolds, 2005, p.38) and interaction (Muijs and Reynolds, 2005, p.43).

Teachers contribute to the 'construction of knowledge' (Mercer, 1995, pp.21-43) through 'guidance strategies' which involve 'intentional, goal-directed ways of talking… which reflect the constraints of the institutional setting'. According to this view teachers support learning by eliciting talk from students, by responding to what the students say and by describing shared classroom experiences (Mercer, 1995, pp.25-26). The techniques for guiding learning formed the three main categories of analysis in the study: Elicitation, Response and Description. These guidance strategies contribute to learning and provide a means of identifying patterns in talk as well as evaluating the contribution the talk makes to learning.

To facilitate analysis of the discourse the study developed a framework for identifying differences in teacher talk, describing how teachers teach and indicating positive or negative effects. During coding of the data the categories of Elicitation, Response and Description were each grouped in two sub-categories, Direct and Cued Elicitations, Evaluation and Correction Responses and Recap and Prospective Descriptions.

Direct Elicitation are questions and Cued Elicitations encourage student talk by providing a clue or prompt that leads to a student utterance, for example a teacher eliciting the word 'pulse' by saying 'you can feel it here' (Mercer, 1995, p.27). In the study the teachers would bring talk out of students using a combination of questions and prompts. If a question did not elicit a good response more clues would be provided to guide the students towards the answer. Clues included starting a sentence for the student to finish, using L1 or inviting peer correction, for example excerpt 1, showing an exchange where the teacher uses a series of prompts to guide the language produced by the student.

Excerpt 1

T: So, let's correct this sentence, OK. Er because it it…

S: (inaudible)

T: The verb is OK, cause, but how to say it…

S: (inaudible)

T: (in L1) it's (in L1) it…

S: it

T: Cause, cause, it causes.

S: causes

T: Causes it causes because it you don't need it. It causes…

S: addiction

Mercer's Response category (1995, pp.32-33) includes confirmation, rejection, repetitions, reformulations or elaborations. In confirmations and rejections a teacher says 'yes' or 'no', 'right' or 'wrong', and repetitions confirm or emphasise the correctness of a student utterance. During coding confirmations, rejections and repetitions were combined into a single sub-category, Evaluation Responses, because all these types of response indicated to the student whether their utterance was correct or not. For example, in one of the lessons the teacher prepares the students to write a formal letter by eliciting forms and content, responding with; 'Yes, yours sincerely' or 'apply, you want to apply for a job, very good'. Reformulations and Elaborations, on the other hand, offer a revised version of what has been said by the student and were combined into one sub-category of Correction Responses, for example; 'Dear, Dear Sir, or if you, if you know the name you can write a name. Dear Sir or Dear madam, comma, yes.'

In the Describe category teachers use 'we' statements, literal recaps and reconstructive recaps (Mercer, 1995, pp.33-41). 'We' statements talk about a past experience that is relevant to the present, for example when a teacher reminds the students about something that happened in the previous lesson. Recaps review aspects of shared knowledge, for example, reminders about previous lessons or drawing conclusions after a discussion or activity. Literal recaps repeat the shared knowledge and reconstructive recaps add further interpretation. These types of talk attempt to connect with experiences that are outside the immediate classroom situation. During coding it proved difficult to distinguish between these guidance strategies, possibly because they apply to a science teaching context, not language teaching. Consideration of the data found the teachers make connections with experiences outside the immediate setting by talking about past and future experiences, so two sub-categories were created, Recap and Prospective Descriptions. Recap Descriptions help the students make connections with previous classroom experiences through explanations of language points or feedback on task performance, and are often given in L1. Prospective Descriptions help the students to connect with activities they are going to do through instructions. Descriptions tend to be longer utterances giving information to the students or reminding the students of something they have done or learned.

3. Results

The talk was coded in the categories and sub-categories and analysed for differences in how the teachers teach in examination lessons. Analysis looked at three levels of talk in the analysis. The first level showed the total amount of talk measured by the number of words used by the teachers. The second level showed the amount of talk in the three categories of Elicitation, Response and Description. The third level showed talk in six sub-categories, Direct and Cued Elicitations, Evaluation and Correction Responses and Recap and Prospective Descriptions. The three levels of talk in the analysis narrow the focus progressively from a broad view of the talk to a detailed view. Categories and sub-categories are shown in Table 2, with the percentage of the total data that they represent. Descriptions made up around half the data, which is not surprising as they tended to be longer utterances giving instructions (Prospective Descriptions) explanations or feedback (both Recap

Descriptions). Similarly Cued Elicitations, which were often single-word prompts, made up the smallest proportion of the data.

Table 2. *Categories and Sub-categories Showing Percentage of Total Data*

Category	**Sub-categories**
Elicitations (21%)	Direct (13%)
	Cued (8%)
Responses (27%)	Evaluation (12%)
	Correction (15%)
Descriptions (52%)	Recaps (24%)
	Prospective descriptions (28%)
All (100%)	All (100%)

Table 3 shows that one of the teachers used significantly more words overall in the examination lessons, but the other did not. This finding can be compared with previous research that found differences in one teacher's examination lessons but not in the other teacher's lessons (Watanabe, 1996).

Table 3. *Words in Examination (E) and Non-examination (N) Lessons for Teachers A and B*

Category	E	N	χ^2	Significant difference
All words teacher A	4997	4874	$\chi^2(1) = 0.77$, $p > .05$	None
All words teacher B	7065	6570	$\chi^2(1) = 8.99$, $p < .05$	Significantly more

The words in the three categories of talk in Table 4 show some highly significant differences in examination lessons for both teachers. The teachers used fewer words for Elicitations and Responses and more words for Descriptions in Examination classes. The teachers appear to be bringing talk out of the students less and responding to their students less in direct interaction and talking more about what the students are going to do or have done. The teachers interact less and talk at length more.

Table 4. *Words in the Categories of Elicitation, Response and Description in Examination (E) and Non-examination (N) Lessons for Teachers A and B*

Category	E	N	χ^2	Significant difference
ELICITATIONS				
Elicitation words Teacher A	1070	1212	$\chi^2(1) = 4.42$, $p <.05$	Significantly fewer
Elicitation words Teacher B	1181	1568	$\chi^2(1) = 27.24$, $p < .001$	Significantly fewer
RESPONSES				
Response words Teacher A	1460	2062	$\chi^2(1) = 51.45$, $p < .05$	Significantly fewer
Response words Teacher B	1096	1583	$\chi^2(1) = 44.26$, $p < .05$	Significantly fewer
DESCRIPTIONS				
Description words Teacher A	2467	1600	$\chi^2(1) = 92.4$, $p < .05$	Significantly more
Description words Teacher B	4788	3419	$\chi^2(1) = 114.2$, $p < .05$	Significantly more

The number of words in the six sub-categories of talk in Table 5 show significant differences in examination lessons for both teachers for Cued Elicitations, Correction Responses and Recap Descriptions, but not for Direct Elicitations, Evaluation Responses and Prospective Descriptions. Both teachers used fewer words for Cued Elicitations and Correction Responses and more words for Recap Descriptions.

Table 5. *Words in the Categories of Direct and Cued Elicitation, Evaluation and Correction Response and Prospective and Recap Description in Examination (E) and Non-examination (N) Lessons for Teachers A and B*

Category	E	N	χ^2	Significant difference
ELICITATIONS				
Direct Elicitation words Teacher A	645	663	$\chi^2(1) = 0.12$ p= >.05	None
Direct Elicitation words Teacher B	810	920	$\chi^2(1) = 3.5$ p= >.05	None
Cued Elicitation words Teacher A	425	549	$\chi^2(1) = 7.89$ p= <.05	Significantly fewer
Cued Elicitation words Teacher B	371	648	$\chi^2(1) = 37.65$ p= <.05	Significantly fewer
RESPONSES				
Evaluation Response words Teacher A	746	764	$\chi^2(1) = 0.1, p > .05$	None
Evaluation Response words Teacher B	547	700	$\chi^2(1) = 9.39, p < .05$	Significantly fewer
Correction Response words Teacher A	714	1298	$\chi^2(1) = 84.8, p < .05$	Significantly fewer
Correction Response words Teacher B	549	883	$\chi^2(1) = 39, p < .05$	Significantly fewer
DESCRIPTIONS				
Prospective description words Teacher A	1162	1093	$\chi^2(1) = 1.1, p > .05$	None
Prospective description words Teacher B	1421	1982	$\chi^2(1) = 46.2, p < .05$	Significantly fewer
Recap description words Teacher A	1305	507	$X^2(1) = 175.72, p < .05$	Significantly more
Recap description words Teacher B	3367	1437	$X^2(1) = 387.7, p < .05$	Significantly more

Both teachers interact less with Cued Elicitations such as prompts, and they provide fewer actual corrections in Correction Responses. Both teachers use significantly more Recap Description words, which refer back to work done, but neither teacher uses significantly more Prospective Descriptions which talk about what the students are going to do. There is no significant difference in Direct Elicitations for either teacher, and Teacher A shows no significant difference in the number of Evaluation Responses. There seems to be no possible examination effect on the teachers' use of questions or their use of indications of whether the student utterance was right or wrong.

4. Discussion

4.1. Does the teaching show evidence of possible washback?

Findings indicate the presence in the data of possible washback on how teachers teach. According to the approach adopted in this study and in a number of previous washback studies (Watanabe, 1996; Alderson & Hamp-Lyons, 1996; Cheng, 1997), significant differences in examination lessons that apply to both teachers may be viewed as possible evidence of washback. Differences that apply to only one teacher may be evidence of influences such as individual teacher preferences. When both teachers use a feature of the

teacher talk more or less washback may be present. In this study both teachers used more Cued Elicitation and Correction Response words and fewer Recap Descriptions, suggesting there may be washback on these aspects of teacher talk.

4.2. What does possible washback show about how teachers teach in examination lessons?

These differences in discourse may represent an examination influence on how teachers teach. Both Cued Elicitations and Correction Responses show teachers guiding students in their learning, whereas Recap Descriptions show teachers telling the students about language or task performance. There seems to be a move away from guiding and towards telling in the discourse of both teachers.

The purpose of *Cued Elicitations* is to guide the students to show a better level of understanding or knowledge. *Cued Elicitations* help students 'take an active part… in the dialogue' and are an alternative to the teacher speaking at length: 'The teacher avoids continuous monologue or the mere provision of missing information.' (Mercer, 1995, p.27). When teachers use *Cued Elicitation* they work with students to build knowledge together. For Edwards and Mercer (1987, p.143) in pupils: 'are being inculcated into what becomes for them a shared discourse with the teacher (discourse in the broadest sense, including concepts and terminology as well as dialogue)… pupils' knowledge is aided and 'scaffolded' by the teacher's questions, clues and prompts to achieve insights that the pupils themselves seem incapable of'. As a result 'effective scaffolding reduces the learner's scope for failure in the task, while encouraging their efforts to advance'. Cued Elicitations, then, perform the function of guiding students to a better knowledge or understanding of an activity or language form. The teachers seem to do less of this guiding in examination lessons.

Correction responses also perform a guiding function in lessons by providing a student with language that improves what the student has just said. Correction responses require engagement from students as they focus on what the teacher wants the students to learn. Correction Responses are focused on teaching language, whereas Evaluation Responses simply inform the students whether they were right or not. The purpose of Correction Responses is to tell students what they should have said in a preceding utterance, as in the excerpt above where the teacher provides the correct form 'causes'. The corrections are intended to help students show their knowledge better. They involve the teacher and student interacting in order to build knowledge. The teachers seem to do less of this knowledge building in examination lessons.

Recap Descriptions provide information to the students about language forms they have used or tasks they have performed. They can refer to language and tasks encountered in previous lessons, servings as reminders of a point previously learned, or they can refer to language and tasks that have just been said or done, serving as feedback or advice. Both teachers use significantly more words on Recap Descriptions in examinations. Much of this talk is feedback about how the students have performed examination tasks.

The field notes and reports showed that examination links are also present in lesson content and methodology. The apparent change in the teacher talk seems to reflect these links. After the observed lessons the teachers stated that, in the examination lessons, their aims were to prepare examination tasks, and an influence on methodology was the need to practise for the examination. Teacher B explains her reasons for choice of methodology thus:

> *The lack of time. That we are very near the exam period. I don't like giving direct explanations, but at this time with these texts whenever something occurs they don't know I tell them, I explain. There isn't time for the method of not explaining directly.*

The teacher reports suggest a shift from correcting to explaining in examination lessons that is linked to the examination itself. Correcting students seems to become less important in the examination lessons, and explaining seems to become more important.

4.3. Do differences in how teachers teach show positive or negative effects?

The data themselves do not indicate whether positive or negative washback is present. Whilst less guiding in the form of fewer words in Cued Elicitations and Correction Responses may suggest a negative influence, with less scaffolding and shared discourse, the increase in Recap Descriptions cannot be said to represent a negative effect, because they also make a positive contribution to learning. The explanations of language points and feedback on student performance in examination tasks may be having a positive effect that counters the effect of less guiding. The category of description used in the study would require development for any claims concerning effectiveness to be made. A further analysis of Recap Descriptions could identify differences in their quality and content, on in the way they make connections with previous experiences, knowledge or learning, for example by looking at the role of explanations about language or feedback about performance.

5. Conclusions

The study found some indications of washback and what it may look like. Washback seems to take the form, on the one hand, of less prompting and correcting talk and, on the other hand, of more utterances that are longer such as explanations and advice. This move away from guiding talk connects with the examination through the lesson aims that are stated by the teachers. The guiding talk is more oriented towards the aims of teaching language forms, whereas the longer utterances focus on the students' examination performance.

This study is comparable with previous washback research, as it identifies any possible washback operating in a limited number of areas. This study goes a little further than previous research by looking in detail at the teacher talk and finding specific aspects of teacher talk that may represent washback. Whilst it may not seem surprising if the approach of a language examination leads teachers to talk more to students about task performance and do less prompting and correction of language forms, previous research has not identified such a washback effect.

If there is a relationship between teacher talk and the examination itself, there are implications for teachers, testers and administrators. Teachers could consider the relative contribution of prompting, correcting and longer utterances to learning for an examination. Testers could consider how they can get teachers to engage with the constructs and purpose of an examination. Administrators could consider the consequences of examinations, if the onset of an examination leads teachers to talk in longer utterances and use less direct interaction to teach language forms,

Three methodological aspects contributed to the study: the study of discourse using transcribed teacher talk, the use of a framework showing pedagogic purpose, and separating how teachers teach from what teachers teach. The combination of attention to classroom discourse related to pedagogic purpose and a focus on how teachers teach as different from what teachers teach and attitudes may support the development of more detailed understanding of washback in further research.

References

Ábrahám, K. & Jilly, V. (1999). The school-leaving examination in Hungary. In Fekete, H., Major, É. & Nikolov, M. (eds.), *English language education in Hungary: A baseline study (pp.21-53).* Budapest: The British Council.

Alderson, J. C. & Hamp-Lyons, L. (1996). TOEFL preparation courses: A study of washback. *Language Testing,* 13, 280-297.

Alderson, J. C. & Szollás, K. (2000). The context: The current school-leaving examination. In Alderson, J. C., Edit, N. and Enikö, Ö. (eds.), *English Language Education in Hungary,* Part II *(pp.9-21).* Budapest: The British Council.

Alderson, J. C. & Wall, D. (1993). Does washback exist? *Applied Linguistics,* 14, 115-129.

Allwright, D. (1991). *The Death of the Method.* Lancaster: Lancaster University, Centre for Research in Language Education.

Allwright, D. & Bailey, K., M (1991). *Focus on the language classroom.* Cambridge: Cambridge University Press.

Andrews, S. (1995). Washback or washout? The relationship between examination reform and curriculum innovation. In D. Nunan, R. Berryand V. Berry, (eds.) *Bringing about change in language education* (pp. 67-81). Hong Kong: Department of Curriculum Studies, University of Hong Kong,.

Andrews, S., Fullilove, J. & Wong, Y. (2002). Targeting washback-a case study. *System,* 30, 207-233.

Anthony, E. M. (1963). Approach, method and technique. *ELT Journal,* 18, 63-67.

Author, (2006). Examination influence on how teachers teach: a study of teacher talk. Unpublished PhD thesis. Lancaster.

Bachman, L. F. & Palmer, A. S. (1996). *Language testing in practice.* Oxford: OUP.

Bailey, K. M. (1996). Working for washback: A review of the washback concept in language testing. *Language Testing,* 13, 257-279.

Burrows, C. (2004). Washback in classroom-based assessment: A study of the washback effect in the Australian adult migrant English programme. In L. Cheng, Y. Watanabe, and With Curtis, A. (eds.), *Washback in language testing: Research contexts and methods* (pp. 3-18). New Jersey: Laurence Earle Brown Associates.

Cadorath, J. & Harris, S. (1998). Unplanned classroom language and teacher training. *ELT Journal,* 52, 188-196.

Chapman, D. W. & Snyder, C. W. J. (2000). Can high-stakes national testing improve instruction: re-examining conventional wisdom. *International Journal of Educational Development,* 20, 457-474.

Cheng, L. (1997). How does washback influence teaching? Implications for Hong Kong. *Language and education,* 11, 38-54.

Cheng, L. (1998). Impact of a public English examination change on students' perceptions and attitudes toward their English learning. *Studies in educational evaluation,* 23, 279-301.

Cheng, L. (1999). Changing assessment: Washback on teacher perceptions and actions. *Teaching and teacher education*, 253-271.

Cullen, R. (1998). Teacher talk and the classroom context. *ELT Journal,* 52, 179-187.

Dinsmore, D. (1985). Waiting for Godot in the EFL classroom. *ELT Journal,* 39, 225-234.

Duff, P. A. (2002). The discursive co-construction of knowledge, identity, and difference: A ethnography of communication in the high school mainstream. *Applied Linguistics,* 23, 289-322.

Edwards, D. & Mercer, N. (1987). *Common knowledge: The development of understanding in the classroom.* Cambridge: Cambridge University Press.

Frederiksen, J. R. (1984). The real test bias: Influences of testing on teaching and learning. *American Psychologist*, 39, 193-202.

Hall, J. K. (1998). Differential teacher attention to student utterances: The construction of different opportunities for learning in the IRF. *Linguistics and education,* 9, 287-311.

Hawkey, R. (2004). *Cambridge ESOL CPE textbook washback study: Full report university of Cambridge ESOL examinations,* Cambridge.

Hayes, B. & Read, J. (2004). IELTS test preparation in New Zealand: Preparing students for the IELTS academic module. L. Cheng, Y. Watanabe and A.Curtis (eds.) (pp. 3- 18). New Jersey: Laurence Earle Brown Associates,

HayMcBer (2000). *Research into teacher effectiveness: A model of teacher effectiveness.* London: Department for Education and Employment.

Hughes, A. (1989). *Testing for language teachers.* Cambridge: CUP.

Kellaghan, T. & Greaney, V. (1992). *Using examinations to improve education: A Study of 14 African Countries.* Washington DC: The World bank.

Kiss-Gulyas, J. (2001). Experiencing the examination design, content, materials and procedures. In J. G.Együd, I. A. Gál and P. Author (eds.) *English Language Education in Hungary, Part IlI: Training teachers for new examinations* (pp. 40-58). Budapest: The British Council.

Kumaravadevilu, B. (1993). Maximizing learning potential in the communicative classroom. *ELTJ,* 47, 12-21.

Lier, L. v. (1988). *The Classroom and the language learner: Ethnography and second-language classroom research.* Harlow: Longman.

Madaus, G. F. (1988). The influence of testing on the curriculum. In L. N. Tanner (Ed.) *Critical issues in curriculum: 87th yearbook of the national society for the study of education,* Chicago, University of Chicago Press.

Mason, J. (1996). *Qualitative researching.* London: Sage.

McDonough, S. (2002). *Applied linguistics in language education.* London: Arnold.

Mercer, N. (1995). *The guided construction of knowledge: Talk amongst teachers and learners.* Clevedon: Multilingual Matters Ltd.

Mercer, N. (2000). *Words and minds: How we use language to think together.* London: Routledge.

Messick, S. (1996). Validity and washback in language testing. *Language Testing,* 13, 241-256.

Mori, J. (2002). Task design, plan, and development of talk-in-interaction: an analysis of a small group activity in a Japanese Language Classroom. *Applied Linguistics,* 23, 289-322.

Muijs, D. & Reynolds, D. (2005). *Effective Teaching: Evidence and Practice.* London: Sage.

Nikolov, M. (1999). Classroom Observation Project. In H.Fekete, E. Major and M. Nikolov (eds.) *English Language Education in Hungary: A Baseline Study* (pp. 221-246). Budapest: The British Council Hungary.

Nunan, D. (1987). Communicative language teaching: Making it work. *ELT Journal,* 41, 136-145.

Ohta, A. S. (1999). Interactional routines and the socialisation of interactional style in adult learners of Japanese. *Journal of Pragmatics,* 31, 1493-1512.

Oskarsson, M. (1973). Comparative method studies in foreign language teaching. *MODERNA SPRÅK,* LXVI, 350-366.

Popham, J. (1987). The merits of measurement driven instruction. *Phi Delta Kappa,* May 1987, 679-682.

Prabhu, N. S. (1990). There is no best method - Why? *TESOL Quarterly,* 24, 161-176.

Prodromou, L. (1995). The backwash effect from testing to teaching. *ELT Journal,* 49, 13-25.

Qi, L. (2005). Has a high-stakes test produced the intended changes? In L.Cheng, Y. Watanabe and A. Curtis (eds.) (pp. 3- 18). New Jersey: Laurence Earle Brown Associates.

Richards, J. C. & Rodgers, T. S. (1986). *Approaches and methods in language teaching: A Description and Analysis.* Cambridge: Cambridge University Press.

Rose, D. (2000). Analysis of Moving Images. In M. W. Bauer and G. Gaskell (eds.) *Qualitative researching with text, image and sound* (pp. 246-262) London: Sage.

Seedhouse, P. (1996). Classroom interaction: possibilities and impossibilities. *ELT Journal,* 50 (1),

16-24.

Shohamy, E., Donitsa-Schmidt, S. & Ferman, I. (1996). Test impact revisited: Washback effect over time. *Language testing,* 13, 298-317.

Sinclair, J. (1982). *The Structure of teacher talk.* Birmingham: University of Birmingham.

Smith, M. L., Edelsky, C., Draper, K., Rottenberg, C. & Cherland, M. (1991). *The role of testing in elementary schools.* Los Angeles, California: Arizona State University

Spratt, M. (2005). Washback and the classroom: The implications for teaching and learning of studies of washback from exams. *Language Teaching Research,* 9, 5-29.

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 and A. Curtis (eds.). New Jersey: Laurence Earle Brown Associates, 3- 18.

Thornbury, S. (1996). Teachers research teacher talk. *ELT Journal,* 50, 279-289.

Vygotsky, L. S. (1962). *Thought and language.* Cambridge: M.I.T. Press, Massachusetts Institute of Technology.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes.* Cambridge: Harvard University Press.

Wall, D. (1999). *The impact of high-stakes examinations on classroom teaching: A case study using insights from testing and innovation theory.* Department of Linguistics and Modern English Language, Lancaster University Unpublished PhD thesis

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: CUP.

Wall, D. & Alderson, J. C. (1993). Examining washback: The Sri Lankan impact study. *Language Testing,* 10, 41-69.

Watanabe, Y. (1996). Does grammar-translation co,me from the entrance examination? Preliminary findings from classroom-based research. *Language testing,* 13, 318-333.

Watanabe, Y. (2004). Methodology in washback studies. In L. Cheng, Y. Watanabe and A. Curtis (eds.) (pp. 3- 18). New Jersey: Laurence Earle Brown Associates.

Wesdorp, H. (1982). Backwash effects of language-testing in primary and secondary education. *Journal of Applied Language Study*, 40-55.

Young, R. F. & Nguyen, H. T. (2002). Modes of meaning in high school science. *Applied Linguistics,* 23, 348-372.