

## WASHBACK OF THE FOREIGN LANGUAGE TEST OF THE STATE EXAMINATIONS IN COLOMBIA: A CASE STUDY

Norma Barletta Manjarrés  
*University of Arizona-Universidad del Norte*

*In the year 2000 a new competence-based State Examination for the Admission into Higher Education was introduced in Colombia. With it, the Foreign Language Test became for the first time an obligatory component, thus officially raising the status of this area of study for the first time. Four years have elapsed since then, so the natural question is whether the test has had any impact on the teaching of English in the country. This aspect of a test, that is, its effect on the teaching and learning, generally known as washback, has been recognized as a very complex phenomenon (Alderson and Wall, 1993; Bailey, 1996; Cheng, 2000; Watanabe, 1997, 2004). The present study aims at describing the washback effect of the English Test in a public high-school classroom in a school in Barranquilla, Colombia. The following data were collected and analyzed for that effect: Official document of the Foreign Language Test, the September 2003 and April 2004 tests, classroom observations, interviews with students and teacher, and the English test used in the class. Though it was not possible to establish that all of what is happening in the context is linked to the introduction of the test, there is a strong correlation between classroom teaching and evaluating practices, and what the examination measures. The paper describes how the test is perceived by the participants, what processes it seems to generate, and also some of the products. Dimensions such as specificity, intensity, and value of the washback are also described. Finally, the paper discusses some of the factors mediating the process of washback being generated.*

### INTRODUCTION

Since the General Education Law was passed in Colombia in 1994, there has been increasing concern about the teaching of foreign languages in the country. The study of a foreign language is perceived as necessary due to the internationalization of the economy, multiculturalism as well as the scientific and technological development of our age. This importance is reflected in the law, which demands that all schools include the teaching of a foreign language in the general syllabus, but it is optional for each institution to decide which language to introduce. However, English is recognized as the most widely used foreign language in the country, so it is the most widely taught.

In the year 2000, the Instituto Colombiano para el Fomento de la Educación (ICFES), the governmental agency responsible for the evaluation of the whole educational system in the country, introduced a new State

Examination for the Admission into Higher Education (Examen de Estado para Ingreso a la Educación Superior). The test responded to the general proposal launched by the Ministry of Education to place the development of competences at the heart of all educational projects and evaluation in the country.

The changes in the examination were prompted by the renewal of educational purposes generated by the General Law of Education, the changes within the disciplines evaluated in the exam in the world context, the introduction of new psychometric models for educational evaluation, and also the research carried out by the ICFES since 1991, with the purpose of evaluating the quality of education in the country. It must also be said that the new examination is a response of the educational authorities of the country to the requirement of the World Bank, which demanded the application of systematic tests in order to grant a loan to the country.

With the new examination, the Foreign Language Test, which had been just an elective since 1980, became an obligatory component of that examination. Official documents defining the general theoretical framework to guide the teaching, learning and evaluation of foreign language competence in the country were issued by the Ministry of Education and the ICFES and made available to all schools and teachers. Acknowledging that many schools have little tradition with the teaching and testing of a foreign language, two years were allowed as a grace period for schools to diagnose the teaching and learning of foreign language competence in each institution, so the first results were not given to the students but only to the schools as feedback. From 2003 on, the general results have been published and all schools, public and private, are being ranked into categories (Very Superior, Superior, High, Middle, Inferior, Very Inferior) according to the results obtained in all the subjects. General results for each subject are summarized for every 'departamento' of the country, but each school has access to its results in each of the tests, including the Foreign Language Test.

The Foreign Language Test was designed to replace the old elective test that was mainly oriented towards the evaluation of grammatical aspects of the foreign language (Instituto Colombiano para el Fomento de la Educación Superior [ICFES], n.d., chap.5) The new test seeks to evaluate the communicative competence of the students, which means "observing those aspects in which they can use the knowledge they possess of that foreign language to act in specific situations which demand their making use of that knowledge" (p. 9, my translation). The document issued by the ICFES as framework for the examination makes explicit mention of the Communicative Language Teaching approach as the official orientation of the teaching of languages in the country based on Littlewood (1981), Widdowson (1978), Brumfit and Johnson (1979) among others. The document also makes mention of the theory of language developed by Halliday (1970), Hymes (1972), and the notion of communicative competence proposed by Canale and Swain (1980), and Canale (1983). Thus the exam relies on a distinction between different levels of competence where knowledge of the grammatical rules and

semantic aspects of the language is meaningful if it is used in specific contexts. Consequently, it is the pragmatic competence that influences all the other levels of competence. Theoretically, the foreign language test sets out to measure how the student can unconsciously use the rules of the language, understand the meanings expressed through the language in different contexts. For evaluative purposes, the document establishes that, though structural, textual, pragmatic, and sociolinguistic elements are all involved in using the language, it is necessary to establish levels of knowledge and appropriation of the language. Thus, the examination relies on Bachman's (1990) inventory of subcompetencies, and distinguishes between *linguistic* competence and *pragmatic* competence. Linguistic competence, which largely corresponds to Bachman's (1990) organizational competence, includes grammatical competence, textual competence and textual coherence. These are the competencies on which all students are evaluated. Pragmatic competence, which accounts for the knowledge of the use of the language and includes illocutive and sociolinguistic subcompetencies, is the focus only for those students that choose to take a foreign language test as part of the flexible component of their exam, that is, the part of the exam that every student decides to be evaluated in more thoroughly.

It was decided then that the obligatory foreign language examination for all the students finishing high-school in the country would test three subcompetencies: *Grammatical competence*, which implies mastering the linguistic code, that is, knowledge of "vocabulary, pronunciation and spelling rules, word formation and function, sentence structures (morphosyntactic aspects), meaning recognition (semantic aspects), intonation (ICFES, n.d., p.28, my translation); *textual competence*, which implies the identification of cohesive elements and the rhetoric organization of a text in which paradigmatic and syntagmatic elements are evident; and *textual coherence*, understood as the ability to structure discourse, logical sequences of ideas, and construct meaning from a text.

The official document acknowledges the lack of audio-oral material as a limitation of the examination. However, "this reason is not sufficient to believe that the test will not give evidence of the communicative competence of the test takers" (ibid, p.24).

The importance of the National Examination cannot be denied: it is a criterion for admission into higher education, it supports the processes of self-evaluation and improvement of educational institutions, and it serves as instrument and basis both for research work and for granting certain educational benefits (ICFES, n.d., chap. 5). There has been intensive work by the Ministry of Education, the local and state level Education offices, as well as non-governmental organizations in large and smaller towns to inform the schools, teachers, students, and the community in general about the new evaluation paradigm and its implications. Therefore, it is natural to suppose that the introduction of the Foreign Language Test has had special repercussion, since this subject is, for the first time in the history of education in the country, considered as important as any of the others. Thus, the

following questions arise : Has the National Examination had any effect on the teaching and learning of English? If so, what kind of effect? How is the examination reflected in the classroom?

In the literature about language testing, the property of the test that concerns its effects on teaching and learning is commonly known as washback. There is already a large body of research done about this specific aspect and it was used to establish the theoretical framework of this study.

## **THEORETICAL FRAMEWORK**

Washback, sometimes referred to as backwash (Biggs, 1995, 1996 in Cheng, 2000), can be generally understood as the effect of an examination on teaching and learning (Chen, 2002; Hughes, 2003; Cheng, 2003). Not all scholars, however, have agreed to its definition. Alderson and Wall (1993) restricted the use of the term ‘washback’ to “classroom behaviors of teachers and learners rather than the nature of printed and other pedagogic material” (p. 118). They would also consider washback to be what teachers and learners do that “they would not necessarily otherwise do” (p. 117). Messick (1996) states that in order to be considered washback, good or bad teaching has to be “evidentially linked to the introduction and use of the test” (p. 16). Moreover, Wall (1997) makes a clear distinction between washback and test impact. The latter would refer to the effect of a test on “individuals, policies or practices, within the classroom, the school, the educational system or society as a whole” (cited in Cheng and Curtis, 2004, p.4). Other researchers (Andrews, Fullilove & Wong, 2002) do not make that distinction and they consider that narrow and wider effects can be included under the term washback. For the purposes of this study, washback will be understood in the wider sense, that is, including what some scholars call ‘impact’.

The concept of washback has been associated with validity. Morrow (1986) refers to “washback validity” to describe the quality of the relationship between testing and teaching and learning” (cited in Cheng, 2000, p.4). For Messick (1996) washback is part of construct validity, and it is an inherent quality of any kind of assessment, especially when the results are used for important decisions. For him, washback contributes to the consequential aspect of construct validity, but information about the operative level of test validity should help one distinguish test washback per se from the effects of good or bad educational practices regardless of the quality of the test. So washback in itself is not a reliable criterion to establish test validity. It is other test properties, like authenticity and directness that are likely to produce washback.

Messick (1996) claims that if a test is deficient because it has construct underrepresentation or construct-irrelevant characteristics, then good teaching cannot be considered an effect of the test, and conversely, if a test is construct-validated, but there is poor teaching, then negative washback cannot be associated with the test. Only valid tests (which minimize construct

underrepresentation and construct irrelevancies), can increase the likelihood of positive washback.

### ***Types of Washback***

Alderson and Wall (1993) distance themselves from a simplistic assumption about the way a test can influence behaviors. Therefore, they developed 15 washback hypotheses according to what is influenced: teaching, learning, content, rate, sequence, degree, depth, attitudes and also the number of teachers or learners affected by a test. Which hypotheses will be put forward depends on the nature of the test, the educational context, and the nature of the decisions that are taken on the basis of the test results. Actually, there seems to be a number of variables in society, education, and schools that determine how washback will appear.

When studying washback, it is also possible to focus on participants (teachers, students, material developers, publishers), process (actions by participants towards learning), and products (what is learned and the quality of learning), as suggested in Hughes's trichotomy model (Hughes, 1993 as cited in Bailey, 1996). Watanabe (2004) proposes disentangling the complexity of washback by conceptualizing it in terms of: Dimension (specificity, intensity, length, intentionality and value of the washback), aspects of learning and teaching that may be influenced by the examination, and the factors mediating the process of washback being generated (test factors, prestige factors, personal factors, macro-context-factors).

Usually researchers focus on one aspect or type of washback. In Alderson and Wall's study in Sri Lanka (Alderson & Wall, 1993; Wall, 1996), the introduction of a test of English as a foreign language proved to produce faster changes in the content of teaching than changes in teaching methodology. Cheng (1997), in the preliminary results of a study of the washback effect of the Hong Kong Certificate of Education Examination in English in Hong Kong secondary schools, reports that washback effect "works quickly and efficiently in bringing about changes in teaching materials [...] and slowly and reluctantly and with difficulties in the methodology teachers employ" (p.1). Cheng introduces the term 'washback intensity' to refer to the "degree of washback effect in an area or a number of areas that an examination affects most" (p.7).

Andrews et al. (2002) found out in their study that the impact of a test can be immediate or delayed. According to these researchers, washback seems to be associated primarily with 'high-stakes' tests, that is, tests used for making important decisions that affect different sectors., for example, determining who receives admission into further education or employment opportunities (Chapman and Snyder, 2000). Madaus (1990 in Shohami, Donitza-Schmitdt & Ferman, 1996) identifies as 'high' such situations when admission, promotion, placement or graduation are dependent on the test.

Another aspect that has been studied is whether the test has been used as a lever for change (Pearson, 1988 in Cheng, 1997), so everything, from textbooks to staff, works to achieve better scores. Cheng (2000) reports on

how tests are often introduced into the education system to improve teaching and learning, especially in centralized countries where tests are considered an efficient tool for introducing changes into an educational system without having to change other educational components (Cheng, 2000). In some countries these tests can be considered “the engine for implementing educational policy” (Petrie, 1987, p.175 in Cheng, 2000, p. 6).

### ***Factors Affecting Washback***

It has been demonstrated that it is simplistic to believe that a test can result in all desired changes in teaching and learning. Education is a complex phenomenon and there are many factors involved in bringing about changes, like the school environment, messages from administration, expectations of teachers and students, for example. Saif (2000) argues that an analysis of the needs and objectives of learners and educational systems should be carried out as a starting point for the research in washback. Wesche (1983 in Bailey, 1996), points out that when tests reflect the situations, content and purpose where learners will use the language, they are likely to improve motivation. Shohamy et al. (1996) consider factors like the status of the subject-matter tested, the nature of the test, and the use to which the test scores are put (p. 300). Wall (1996) provides a list of factors which might have prevented the examination in Sri Lanka from providing an effective ‘lever for change’. These ranged from teachers’ factors (lack of understanding of exam, the nature of the change desired, resistance to change, unfamiliarity with the test format and content) to more macro factors like gap between designers of test and teachers, lack of well trained teachers, overload of teachers, etc. Besides, according to Andrews et al. (2002), the innovating effect of a testing innovation is affected by the teachers and how they interpret the innovation, which may differ from what the conceivers of the test had in mind. Another variable can be the published materials in use (Andrews et al., 2002).

As can be seen, washback is a very complex notion. It can refer to the effect of an examination in the classroom, but also in the school, in the educational system and also in the society. Besides, this effect does not always take place directly but it is mediated by a number of factors, like the teachers’ perception of the test, the status of the test as well as that of the subject – matter tested, the macro – context where the examination is used, the purpose of learning the language in the context, among others. Additionally, in order to study the washback effect, it is necessary to look at the people that participate in the educational process, to the actual classroom events and activities, and to the outcomes of these processes.

## **THE STUDY**

The general objective of the present study is to describe the washback effect of the English Language Test of the State Examination in the teaching of English in a 10<sup>th</sup> grade classroom at a public school of long tradition in Barranquilla, Colombia. This school, which conflates middle and

high-school grades, has about 800 female students from lower socio-economic status. Its academic standards are not particularly high among the schools of the town according to the results of the State Examination, but it certainly qualifies as one of the best among the public schools in Barranquilla, where private, more affluent institutions usually rank higher according to the results of the examination made public to the general public (Cámara de Comercio, 2001).

To study the washback phenomenon it was necessary to start by analyzing the official document issued by the ICFES as framework for the Foreign Language Tests as well as the actual tests that students took in 2003 and 2004. This information was then compared with the classroom practices recorded from ethnographic observations (5 lessons were observed in the development of a complete unit), an interview with three students, a formal interview with the English teacher, and informal interview with the latter together with another English teacher of the school. A class examination done in class was also collected and analyzed. The observations and the interviews were fully transcribed and some parts of them were translated into English by the author of this paper.

For the purpose of the analysis, the trichotomy proposed by Hughes (1993 cited in Bailey, 1996) was found useful in order to distinguish between the different instances affected by the test: participants, processes and outcomes. In the discussion section the framework suggested by Watanabe (2004) was followed to account for a number of dimensions of the washback: intensity, intentionality and value, as well as for the analysis of the factors that mediate washback.

## **DATA ANALYSIS**

### ***The Foreign Language Test***

An insight into the test was obtained through an analysis of the official document of the Foreign Language Test of the ICFES examination, available for teachers and interested people through the web site of the ICFES. It contains the general theoretical frame of the teaching of foreign languages in Colombia. A review of the legal dispositions makes it clear that since the General Law of Education was issued in 1994, all schools have to teach at least one foreign language and that this teaching has to aim at developing communicative competence. Special attention is given to Canale and Swain's (1980) and Canale (1983) definition of communicative competence with the distinction between grammatical, sociolinguistic, discursive and strategic competences. However, the concept of communicative competence was made operative through the notions of linguistic and pragmatic competences (Bachman, 1996). However, only linguistic competence with its three subcompetencies - grammatical competence, textual competence and textual coherence - is included in the obligatory test all the students have to take.

The test includes 35 multiple-choice questions, distributed in seven different types of tasks or 'contexts of evaluation'. Table 1 summarizes the

types of tasks or “contextos de evaluación” with their distribution in the 2003 and 2004 versions of the test, as well as the type of subcompetence evaluated in type of task.

Table 1: English Test Item Types

<b>Contexts of Evaluation (Tasks)</b>	<b># of Items</b>	<b>Competence</b>
Identification of grammatically correct construction in a given context.	5	Grammatical
Interpretation of graphics.	5	Grammatical
Cloze test where missing words are verbs, prepositions, connectors, adjectives, nouns, etc.	6	Textual
Organization of paragraphs.	2	Textual
Dialogue completion (students identify the missing turn).	3	Textual coherence
Situation comprehension	2	Textual coherence
Text comprehension	12	Textual/Textual coherence.

Out of the 35 items of the test, ten test exclusively grammatical competence: knowledge of discrete vocabulary items or grammar structure. The other 25 items test textual competence or textual coherence. The graphics are occasionally confusing with more than one interpretation possible. The cloze test requires students to know the correct verb form or the right preposition in a certain sentence. In some cases it also requires understanding of relationships between adjacent sentences or clauses, asking students to choose the correct conjunction, connector or to interpret a reference. The dialogue completion task is not authentic in that the test takers have to read the line after the blank of the missing turn, in order to get the answer right. This is never a requirement for using the language in real life. Besides, sometimes more than one answer is possible. The ‘organizing paragraphs’ items, another non-authentic task, actually asks students to order sentences or chunks of sentences in a logical order. It sometimes seems to be testing general logical abilities and one may feel like solving a puzzle, rather than be tested on actual knowledge of how to connect pieces of discourse or use discourse signals. The ‘text comprehension’ items include simplified texts (short sentences, simplified vocabulary and structures). These items contain questions that aim at testing global understanding (the general idea of the text, the intention of the text, inferences that can be drawn from the text) but also discrete vocabulary items, paraphrasing of phrases and local reference interpretation. In general, the questions in the examination range from those that test basic knowledge of vocabulary and grammar to those that ask for general understanding of short



simplified texts, as well as inference making of specific passages. The test still has a considerable focus on form. Sometimes it is possible to answer a question without fully understanding the meaning or the context where the language is used. Some items ask for understanding beyond sentence level, but the larger discourse is never longer than one paragraph of non-authentic language. Pragmatic competence is not tested.

### ***Participants: The Teacher***

Among the factors that can mediate the washback effect is the teacher (Wall, 1996) and her/his perceptions about the examination, its nature, purposes, relevance in the context, etc. These aspects were explored in an interview with the 10<sup>th</sup> grade English teacher. (See the questions asked in the interview in Appendix 1). She is a woman with many years of teaching experience in the school and has attended a number of in-service training workshops, including a visit to the USA to get training in the use of a software for the teaching of English. In the interview she was together with a colleague. Both expressed a certain degree of satisfaction with the fact that a formal English department had been constituted in the past year, thus formalizing the importance of English within the institution. Consequently, only English teachers can now teach the subject, while in the past Spanish teachers would do it to complete their academic load. Besides, following the suggestion made by the English teachers, the administration agreed to divide each class into two groups for the English lessons in order to have groups of about 20 students each.

When asked about the goals of the English classes, the teacher said that the English department wants students to be able to have a conversation in English, to write compositions, to understand when they listen to somebody, and also to read and understand. However, they also attach special importance to grammar: “Some people think that speaking is the most important, but they do not realize that if they have good grammar basis, this oral part is easy.... One thing is not separated from the other (my translation)”.

The syllabus was restructured that year based on a commercial textbook because, according to the teachers, it develops the four communicative skills and includes topics the students like. However, when she explained grammar, she resorted to other books she knows. When asked specifically how the communicative approach is reflected in the textbook, she said it is reflected in that it entails many activities to develop oral skills and also the graphs, which can be used for the students to speak and explain.

Communicative competence for her is “those abilities that students can develop taking into account good listening, reading, writing and speaking in a foreign language (my translation)”. They (English teachers) do not usually do oral evaluations because it is necessary to have written proof of their students’ performance and the written tests can be sent home for the parents’ signature. Besides, a written test “makes students study grammar.”

The issue of the State Examination was brought in by her during the interview when asked what she thought about the teaching of English. She

said: “We have to consider the ICFES examination very, very, very, very, very much (my translation)”. She pointed out that the speaking and listening skills are not evaluated there, so she wonders if the aim of the test is reading comprehension or grammar. She recognized that “everybody is obsessed” with the exam.

### ***Participants: The Students***

The interview done with three of the students of the class observed dealt with their perceptions about the importance of English, the classes, and their competences. In their answers, students reflected general awareness of the instrumental benefits of knowing English and the importance of the ICFES English Test. They say they want “to go deeper in English because it is important for the ICFES” (my translation). Outside school they use English when they watch movies since they listen and read the captions. Occasionally they have fun speaking in English when they can. They think of chatting as a potential use of English.

They have taken two preparatory tests for the ICFES examination “to reinforce their knowledge, and to know how much they know of English.” They have the perception that they did well. In general they have the perception that what they have gained most in is grammar and pronunciation.

### ***Processes***

Processes, according to Hughes (1993 in Bailey, 1996), refers to material development, syllabus design, changes in teaching methodology, and testing strategies among others.

Just as the teacher stated in the interview, the syllabus of the course was textbook driven, though sometimes she would not follow the exact order or would choose to skip some of the activities. During the observations, the book was followed while it dealt with discrete points, like quantifiers, vocabulary and guided writing. More extensive writing and project development was not dealt with in class, but assigned as homework before the topic was actually developed.

An analysis of the classroom observations showed that from the seventeen activities developed during the time observed, nine of them were directed towards the development of grammatical competence (gap filling exercises, graph interpretation, translation of discrete items, and pronunciation practice). Beyond that was a cloze exercise (where learners had to fill in the appropriate quantifier) and two guided writing tasks where students were asked to write down a sequence of four events using sequence connectors. Other activities, like oral presentations and dialogue dramatizations, which are potentially suitable for developing other competencies, were not exploited as such because they turned out to be rote memorization of written texts (or dialogues), performed with poor intonation and with the only objective of getting credit for doing it. In other cases the presentations consisted in reading aloud texts that contained a number of mistakes. There was no evidence of the

other students' following or understanding the presentations or dialogues. The researcher herself found it difficult to understand them.

The students were most of the time practicing rules in artificial exercises that focused on form, usually at sentence level, with predictable responses. Feedback on presentations and dialogues was very little, usually very general. For grammar exercises, the correct answer was supplied and sometimes also the repetition of the rule if there were errors.

During the observation time a test was administered. Table 2 summarizes it and shows that in many ways it resembles the English Test from the ICFES examination.

Table 2: In-Class Written Examination

Type of item/Instruction	Competence	Example	% of correct answers in the class.
Matching	Textual	How much oil do you need? How many tomatoes do we need? a. Only a few b. Only a little	68
Fill in the blank	Grammatical	Are there ____ eggs? I don't a much meal, but I eat ____ fish.	39
Dialogue completion. Using how much/ow many/ a few, a little or a lot of.	Textual	1. How much water does Betty drink each day? _____ 2. _____ ? She can buy a lot of apples	31
Writing. Describe recipe using sequence words	Textual coherence	Students had to use sequence words: after that, finally, first, next, then. Apart from this, only blank lines were given.	42
Fill in blanks in a dialogue with the correct future form.	Grammar	-Hi, Mike. What's new? -I'm making plans for the weekend (I, go) _____ to the Latin Music Fiesta on Friday.	10
Average			38

### ***Products***

The products in a washback study refer mainly to what is learned. In this case, the focus is the development of communicative competence. As can be seen from the results of the in-class test (see Table 2), and from the analysis of the performance of the students during the class observations, students seem

to be working exclusively towards the development of linguistic competence in class. However, they obtained better results in those items that asked to demonstrate textual competence (68%) than in those items that evaluated grammatical competence (10% and 39% for the two items that tested it). That is, they are more competent dealing with connected texts of two or more sentences, rather than with discrete grammatical items. In terms of the ICFES examination, scores within the range of 31 and 45 in a 100-point scale would be considered low-mid, while scores within a range of 46 and 70 would be considered high-mid. Altogether, they would be considered low-mid, because the average score of the class was 38%. Besides, there is no trace of the development of any other subcompetence within the larger notion of communicative competence.

## DISCUSSION

The central question of this study was whether the English Test has had any washback effect on the teaching of English in the specific context of this study, which can be considered a representative case of public schools in big towns in the northern part of Colombia, not particularly privileged, but certainly not especially poor or disadvantaged. The results of the study do not allow us to give a unique unequivocal answer as to what is the effect of the English Test in the context of the study because it is not possible to establish with certainty how was English taught and learnt in the context before the ICFES examination was introduced. However, from the analyzed data it is possible to document certain trends that seem to go in line with the test, but which are of different nature and strength. Such dimensions as specificity, intensity and value of the washback (Watanabe, 2004) will serve to better characterize the type of effect the examination is producing in the context of the study. These dimensions will be treated separately for the sake of this report, but as will be seen they all interact with each other.

### *Specificity*

According to this dimension, the washback of a test can be general or specific. General washback refers to an effect related to tests in general, while specific washback refers to washback associated with particular feature of the test. The ICFES test was meant as a step in a lengthy process of improving the quality of teaching in the country and as a measure to quantify the advances made by the schools since the introduction of English as an obligatory subject. Logistically and administratively the school is making an effort providing more teachers and classrooms for the teaching of English in order to have fewer students in each class, and allowing only qualified teachers to teach the subject. Besides, students take preparatory tests in the school in order to score higher in the exam. A textbook providing ample practice in linguistic competence was introduced and most classroom time is devoted to this aspect as a result of a certain curriculum alignment with the test. As a result, the competencies that are being developed are to a large extent the ones tested in

the ICFES exam. Though no pressure, panic or urgency is felt inside the classroom to work towards better performance in the examination (it was never mentioned), teacher and students are well aware of the importance and the demands of the examination and, in a way, they admit that what they do is related to this challenge, be it only for the sake of not doing that bad. The teachers feel they are evaluated through this, but students seem to associate the exam with a wider general goal of knowing English. Consequently, it can be said that the test has contributed to strengthening the perception of the importance of English inside the school and it is related to perceptible moves inside the institution towards facing the challenge. It could be argued that if there were no exam, the changes would hardly have taken place. After all, the State Examination is a very strong justification for budget allocation. Consequently, it can be said that the test has had a general positive washback, because it has prompted a number of changes towards improving the quality of teaching, but most of this effect cannot be directly associated to any specific feature of the test, just with the fact that *there is* such a test.

### ***Intensity***

The intensity of the washback, according to Watanabe (2004), can be described as strong or weak. While in our case it can be said that the test has had noticeable and a rather strong impact at administrative, logistic and planning level, the same is not felt inside the classroom. The data do not show that the English Test determines everything that happens there or that it is substantially different from what was done before. Unfortunately, there is no reliable data as to how English was taught before the test.

According to Watanabe (2004), the intensity of the washback could be a function of the importance of the test's consequences: the higher the stakes, the stronger the intensity. In this study, the exam can nominally be considered a 'high-stakes' test, but it works differently for each of the participants. Actually, in the context of the study, as is the case with most public schools, not all the students would try to enter a higher-education institution. For most of them, the examination will simply be just a formal requirement for receiving the high-school completion certificate.

The school as an institution sees the ICFES test as the instrument whereby it is ranked among all the other schools. The teachers are also evaluated by the results of the test because it would be particularly shameful for them if the scores in their area are lower than those obtained in other areas. However, the degree of institutional or faculty effort is mediated by what the school considers is the expected performance of the school, and what the teachers consider the general performance of the subject should be as compared with other subjects. So, each institution and teacher will react accordingly. In our case, however, we could see that the test has had strong washback for school and teacher.

Now, students in that particular school may not consider the stakes of the test that high, given that they might not be expecting to get into higher education, so the test would not have any practical consequence for them. So

here, mainly private goals, personal commitment to school related responsibilities, particular interest in the subject, or even the vague idea or chance that there is a possibility of needing English language skills in the future, mediate the effect of the test.

### **Value**

In order to judge whether the value of the washback is positive or negative, we need to follow Alderson's (1992 in Watanabe, 2004) suggestion of identifying the audience, that is, the effect of the test can be evaluated differently according to the parties involved. On one end, the ICFES needs the results of the test for the sake of measuring improvement of education and responding to the demands of the World Bank, so the exam serves that end, and apparently positive changes are being implemented in schools. On the other end are the students and what and how much they learn. A special reflection is due here.

The general proposal of the test in the theoretical discussion of the official document states that the exam is meant to have students show their ability to go beyond applying the rules governing the language. It is supposed to give evidence of test takers' use of the language for communicative purposes, creative use of the language, and recognition of the sociolinguistic levels of the users of the language. However, the tests analyzed are limited to the lower or more restricted competencies. Pragmatic subcompetencies are not included in the general test and that certainly has had its effect. The teacher has the general impression that the test either tests grammar or reading comprehension. This is strongly reflected in the classroom activities. Her words reflect the same kind of ambivalence of the theoretical framework of the test: She wants her students to be able to talk, to understand when talked to, to write and to understand written texts, but her teaching does not include these abilities as such. When students are asked to speak, they read or recite texts. She expects them to memorize texts, so that later they can use these chunks when needed. The focus in the class was mainly on forms, not on communicating meaning. Students perceive writing or speaking almost as a luxury, because if they engage in them very much, they would 'lag behind'. Besides, the feedback they receive on the few tasks that might have been communicative (presentations, writing recipes) was minimal, and most attention was paid to grammatical and pronunciation accuracy. The findings seem to confirm Messick's (1996) claim that washback is related to the construct validity of the test. There is obvious underrepresentation of the concept of communicative competence in the test and indirectly then in the classroom; thus, the washback can hardly be positive.

However, other factors could be argued to influence the washback. In the particular case of the study, the teacher was evidently grammar-oriented. Her personal beliefs about what is important obviously counted at the time of making decisions about textbooks, activities, types of interactions favored, feedback and formal evaluations, etc. Especially significant for her seemed to be the belief that students cannot cope with communicative activities if there is

not proper grammar knowledge as a basis. And this is especially negative because the data show that her students were developing textual competence at a higher level than they were developing grammatical competence (see Table 2 in-class test results) and she might decide it is not worth continuing with the development of other competencies until they do better at the grammatical level.

Negative as these outcomes seem to be, they still do not say much about what students can really do with the language, or whether they would be able to use it for authentic purposes. This aspect has not been considered either in the classroom practices or in the in-class evaluation.

Table 3 summarizes a model of how washback works in the context and the type of washback that the different factors seem to be generating. The test produced general awareness of the importance of English, reduced class size and seems to contribute to the generation of ideal goals in line with the communicative competence construct. These are in themselves part of the general positive washback effect, which was perceived here as 'strong' and 'positive'. However, since it is 'general' washback, as a factor for the outcome of the test in terms of learning, its effect has shown to be rather weak. What seems to be crucial is the teacher and the decisions she makes (syllabus, activities, evaluation, etc.). These decisions, however, cannot evidentially be linked to the examination because nothing in the class or in the interviews can uncontroversially show such a direct relation, but there appears to be, nevertheless, a strong correlation.

Table 3: How Washback Works

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

## CONCLUSIONS AND RECOMMENDATIONS

This study shows the complexity of the washback phenomenon and the appropriateness of qualitative research methods to investigate it. It has documented an overall positive washback effect of the introduction of the foreign language test in that there is general awareness of the importance of improving the teaching and learning English as a foreign language in the school of the study. Concrete steps have been taken to help achieve that goal, for example, the reduction of the class size, the requirement of due qualifications to teach English as well as the change of textbook.

The communicative competence is reduced in the ICFES English to three subcompetencies - grammatical competence, textual competence, and textual coherence. The orientation of the classroom practices, including class tests, strongly correlate with the direction set by the exam. The scope of both the test and the class goals is quite limiting, because most of the time the students are asked to focus on linguistic features, and any activity that goes beyond this (dialogues, writing, and presentations) is seen as a waste of time. Students spend most of the time solving grammatical exercises from the book or the board and translating unknown words from Spanish into English or vice-versa. Not surprisingly, the students' level of linguistic competence is rather low.

It is hoped that in the very near future the educational authorities decide to include the evaluation of pragmatic competence in the test, so that it gains face and construct validity and teachers make a definite step toward focusing beyond linguistic forms into meaning negotiation. As it is, the test does not fulfill its theoretical goal of testing students' ability to use the language creatively for their communicative purposes, or to recognize the sociolinguistic level of the users of the language. It is also essential that the teacher gets extensive professional development opportunities in relation to what communicative competence is and how it can be developed and tested. It is also hoped that more direct and authentic language and tasks are included in the test. If students are not asked to speak or write beyond recitation or mere copying, those practices are unlikely to be felt as relevant and important in the classrooms. It also seems essential that an analysis of the needs of English in our context be carried out, so that teaching and testing match these needs and students feel a real necessity to learn the language and do not see the foreign language test as a mere formality for finishing high-school.

The English Test in its current form can also be seen as an initial step towards a long-term goal of raising the standards of communicative competence in the country, in which case the context under study seems to be VERY slowly going in the right direction. However, there is no evidence available in this study showing that what is being done currently is qualitatively different from what had been done before the exam was introduced. It is important then to keep track of future changes in the test and of the ways in which the school and teacher react to them, for which a longitudinal study will be required.



## REFERENCES

- Alderson, J.C. (1992). Guidelines for the evaluation of language education. In J.C. Alderson & A. Beretta (Eds.), *Evaluating second language education* (pp. 274-304). Cambridge, England: Cambridge University Press.
- Alderson, J.C., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14, 115-129.
- Andrews, S., Fullilove, J. & Wong, Y. (2002). Targeting washback – a case study. *System*, 30, 207-223.
- Bachman, L. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L. & Palmer A. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Bailey, D.M. (1996). Working for washback: A review of the washback concept in language testing. *Language Testing*, 13, 257-279.
- Brumfit, C., & Johnson, K. (1979). *The communicative approach to language teaching*. Oxford: Oxford University Press.
- Biggs, J. (1995). Assumptions underlying new approaches to educational assessment. *Curriculum Forum*, 4 (2), 1-22.
- Biggs, J. (1996). *Testing: To educate or to select? Education in Hong Kong at the cross-roads*. Hong Kong: Hong Kong Educational Publishing Co.
- Cámara de Comercio. (2001). La Revista de la Cámara de Comercio (Circulates with *El Herald*), 2, July 30.
- Canale, M. (1983). From communicative competence to communicative language pedagogy. In J. Richards & R. Schmidt. *Language and Communication* (pp. 2-27). London: Longman.
- Canale, M. & Swain, M. (1980). Theoretical bases to communicative approaches to communicative teaching and testing. *Applied Linguistics*, 1, 1-47.
- Chapman, D. & Snyder, C. (2000). Can high stakes national testing improve instruction: Reexamining conventional wisdom. *International Journal of Educational Development*, 20, 457-474.
- Chen, L. (2002). *Washback of a public exam on English teaching*. (ERIC Document Reproduction Service No. ED472167).
- Cheng, L. (1997). Does Washback influence teaching? Implications for Hong Kong. *Language and Education*, 11 (1), 38-54.
- Cheng, L. (2000). *Washback or backwash: A review of the impact of testing on teaching and learning*. (ERIC Opinion papers ED 442280).
- Cheng, L. (2003). *Looking at the impact of a public examination change on secondary classroom teaching: A Hong Kong case study*. *Journal of Classroom Interaction* 38 (1), 1-10.

- Cheng, L. & Curtis, A. (2004). Washback or backwash: A review of the impact of testing on teaching and learning. In L. Cheng, Y. Watanabe & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods*. (pp. 3-17). Mahwah, N.J.: Lawrence Erlbaum.
- Cheng, L., Watanabe, Y. & Curtis, A. (Eds.) (2004). *Washback in language testing: Research contexts and methods*. Mahwah, N.J.: London: Lawrence Erlbaum.
- Haliday, M.A.K. (1970). Language structure and language function. In J. Lyons (Ed.), *New horizons in linguistics*. Harmondsworth. UK: Penguin.
- Hughes, A. (1993). Backwash and TOEFL 2000. Unpublished manuscript, University of Reading in England.
- Hughes, A. (2003). *Testing for language teachers*. (2<sup>nd</sup> Ed.) Cambridge: Cambridge University Press.
- Hymes, D. (1972). On communicative competence. In J. P. Pride & J. Holmes. (Eds.) *Sociolinguistics: Selected readings*. Harmondsworth, UK: Penguin.
- Instituto Colombiano para el Fomento de la Educación Superior. (n.d). Nuevo examen de estado para el ingreso a la educación superior: Cambios para el siglo XXI- Idiomas. Retrieved November 1, 2004 from <http://www.icfes.gov.co/>
- Littlewood, W. (1981). *Communicative language teaching*. Cambridge: Cambridge University Press.
- Madaus, G. (1990, December). *Testing as a social technology*. Paper presented at the inaugural annual Boisi Lecture in Education and Public Policy, Boston College.
- Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13 (3), 241-256.
- Morrow, K. (1986). The evaluation of tests of communicative performance. In M. Portal (Ed.), *Innovations in language testing*. (pp. 1-13) London: NFER/Nelson.
- Pearson, G. (1988). Tests as levers for change. In D. Chamberlain & R. J. Baumgardner (Eds.) *ESP in the classroom practice and evaluation* (pp. 98-107). Modern English Publications.
- Petrie, H. G. (1987). Introduction to evaluation and testing. *Educational Policy*, 1, 175-180.
- Saif, S. (2000). Theoretical and empirical considerations in investigating washback: A study of ESL/EFL learners [Abstract]. *DAI*, 61, 03, 966.
- Shohamy, E., Donitsa-Schmidt, S. & Ferman, I. (1996). Test impact revisited: Washback effect over time. *Language Testing*, 13 (3), 298-317.
- Wall, D. (1996). Introducing new tests into traditional systems: Insights from general education and from innovation theory. *Language Testing*, 13 (3), 334-354.

- Wall, D. (1997). Impact and washback in language testing. In C. Clapham & D. Corson (Eds.), *Encyclopedia of language and education: Vol. 7. Language testing and assessment* (pp. 291-302). Dordrecht: Kluwer Academic.
- Watanabe, Y. (1997). Constructing a classroom observation scheme for the test impact research. *Sophia Linguistica*, 41, 297-313.
- Watanabe, Y. (2004). Methodology in washback studies. In L. Cheng, Y. Watanabe & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods*. (pp.19-36). Mahwah, N.J.: Lawrence Erlbaum.
- Wesche, M.B. (1983). Communicative testing in a second language. *Modern Language Journal*, 67, 41-55.
- Widdowson, H. G. (1978). *Teaching language as communication*. Oxford: Oxford University Press.

## APPENDIX 1

### Guiding Questions For The Interview With The English Teacher.

1. How would you define communicative competence?
2. What evaluation activities do you use? With what purpose?
3. What are the abilities your students are developing in the English class? Mention them in order of importance.
4. What level of competence, or what do students need to achieve to be promoted to the next grade?
5. What do you do in class to help them develop the required level of competence?
6. What aspects of the general guidelines developed by the ministry of education have you used for the design of the program and the activities in class?
7. What has been the influence of the ICFES examination in the planning and implementation of the program and the activities in class?