

## WASHBACK OF A STANDARDIZED TEST ON SECONDARY SCHOOL EFL TEACHERS

RODRIGO ESTEBAN MUÑOZ ALVARADO  
Universidad San Sebastián – Chile\*

VALERIA SOLEDAD MUÑOZ RUSSELL  
Universidad San Sebastián – Chile\*

DRAGO PATRICIO VRSALOVIC MIHOEVIC  
Universidad San Sebastián – Chile\*

**ABSTRACT:** The purpose of this research was to determine the washback caused by a standardized test of English on eleventh year students. In order to obtain the data, 22 EFL teachers from Puerto Montt in Southern Chile, who were teaching English in eleventh grade at government subsidized private schools, answered a seventeen-question long online questionnaire three months before the test took place. The data was analyzed and processed and the results evidenced a tendency towards a low impact from the test. Nevertheless, teachers perceived that once the results of the test become available they would be used by students, parents and school authorities to judge their teaching practices. These results seem to contradict the specialized literature regarding the washback caused by standardized examinations implemented on a national scale; however, the circumstances under which this test was implemented may provide the reason for this apparent discrepancy.

**KEYWORDS:** EFL, Standardized test, Washback.

### *IMPACTO DE PRUEBA EN PROFESORES DE INGLÉS COMO IDIOMA EXTRANJERO DE UN EXAMEN ESTANDARIZADO APLICADO A ESCALA NACIONAL*

*RESUMEN:* El propósito de este trabajo fue determinar el impacto de prueba (washback) provocado por un examen de inglés estandarizado aplicado a estudiantes de tercero medio por el gobierno de Chile. Con el fin de obtener los datos, 22 profesores de inglés como idioma extranjero de la ciudad de Puerto Montt que se encontraban enseñando dicha asignatura en terceros medios en colegios particulares subvencionados respondieron un cuestionario online de 17 preguntas tres meses antes de la aplicación del examen. Los datos fueron analizados y procesados y los

\* Para correspondencia, dirigirse a Rodrigo Muñoz (rodrigomunoz.1983@gmail.com), Valeria Muñoz (valeria.munoz@uss.cl) y Drago Vrsalovic (dragov42@gmail.com).

*resultados mostraron una tendencia hacia un impacto de prueba bajo. No obstante, los docentes percibieron que una vez los resultados del examen se hicieran públicos serían utilizados por los estudiantes, apoderados y directivos escolares para juzgar sus prácticas pedagógicas. Estos resultados parecen contradecir la literatura especializada respecto del impacto de prueba causado por exámenes estandarizados aplicados a escala nacional. Sin embargo, las circunstancias bajo las que este examen fue implementado pudiesen explicar la razón por esta aparente discrepancia.*

*PALABRAS CLAVE: Inglés como idioma extranjero, Prueba estandarizada, Washback.*

## INTRODUCTION

The first half of the 2010 decade witnessed an increase in the number of standardized tests that were used to measure learning in Chilean schools. Apart from the traditional reading, math, science and social studies tests that had been used for nearly two decades, authorities added tests on writing, physical education, ICTs and English. Throughout the last eight years, some of these tests have been discarded. However, the English examination seems to be placed in high regard. In fact, educational authorities have revealed their intentions of adding an English part to the battery of the national university entrance exams.

Studies related to the application of standardized tests have shown that such policies most of the times tend to do more harm than good as these tests add unnecessary pressure on teachers and students. Learning is sometimes left aside as training on test-taking strategies becomes the priority. The effect of tests on teaching is referred to as *washback*.

This article addresses the 2017 application of a standardized test of English to assess students' learning of the foreign language placing emphasis on the impact of said examination on teachers' practices.

## LITERATURE REVIEW

### *Washback*

In simple words, *washback* refers to the influence that tests have on teaching which results in a set of actions that probably would not occur under different circumstances (Alderson & Wall, 1993). Through the last several years, such influence has also been labeled *Backwash*, *Test impact*, *Systemic validity*, *Consequential validity*, and *Curriculum alignment* (Pan, 2009). For the purpose of this study, *washback* will refer to the influence that a high-stakes test (English SIMCE) has on teachers' decisions regarding their teaching as a whole, i.e. contents, methodology, assessment, among others as well as the decisions made by school authorities in regards to the way English should be taught and emphasized within their schools.

According to Pan (2009), the aforementioned influence might be positive or negative, hence the terms *positive washback* and *negative washback*. The characteristics

of positive washback can be listed as but are not limited to the creation of beneficial teaching-learning activities (Brown, 2004), motivation to fulfill learning and teaching goals (Anderson & Wall, 1993), the design of a new syllabus or the alteration of the old ones (Cheng, Watanabe & Curtis, 2008). On the contrary, negative washback may arise in the form of the narrowing of the contents to be taught (Shohamy, 2005), ignoring activities that are not directly related to the test (Pan, 2009), teachers' anxiety and stress, especially if test results are to be made available to the public (Fish, 1988), and learning test-taking skills instead of using the language for communicative purposes (Pan, 2009).

Washback arising from high and low-stakes tests has been extensively documented (Cheng, Watanabe & Curtis, 2008; Muñoz, 2017). Fish (1988) states that the publication of results drawn by high-stakes tests caused extra accountability pressure and stress on teachers. Additionally, since it is likely to improve test scores (at the expense of learning), teachers may feel tempted to practice test-taking skills and drilling. Similarly, teachers may feel discouraged to try methods and teach contents that are not compatible with standardized tests (Noble & Smith, 1994). Cheng (2008) states that the effects of high-stakes tests on teaching and learning are "powerful." However, she also indicates that these effects are seen more prominently in the selection of content rather than the methodology used by teachers. A similar assertion can be found in Alderson and Wall (in Cheng, Watanabe & Curtis, 2008), although they also add test design as another element clearly influenced by the insertion of a high-stakes test. Both cases seem to support Cheng's perspective in the sense that tests cause quick changes in what is taught rather than how it is taught. Along the same line, Qi (2008) proved that after 15 years of an annual implementation of a high-stakes English test in China, instructors preferred to teach language as discrete points, rather than communicatively, and there was an emphasis on the skills tested in the national examination, although the test constructors' intentions had been to promote positive curricular innovation.

### *Chilean English SIMCE*

The SIMCE (Sistema de Medición de la Calidad de la Educación) tests were established in Chile in 1988. At first, they were in the hands of Pontificia Universidad Católica and later in 1992 the Chilean State took over their implementation. However, the SIMCE was not the first instance in which a standardized test was used in order to measure Chilean students' learning and/or gather data to aid in the decision making processes. In fact, two different instances with the same purposes had been put into practice beforehand: first, Prueba Nacional in 1968 and then Programa de Evaluación del Rendimiento in 1982 (Ortiz, 2010). Through its lifespan the set of exams that compose the SIMCE have varied greatly in terms of the subjects that are assessed as well as the school years in which each area is tested.

For the purpose of this study, the most relevant change in the SIMCE was announced in the yearly May 21 Presidential speech in 2010. Then President, Sebastián

Piñera informed that an English test would be added to the SIMCE set. Days later, then Minister of Education, Joaquín Lavín, revealed further details on the test indicating that its design had been carried out by Educational Testing Service (ETS), that it would assess listening and reading comprehension and that students who pass it would receive an official certificate (“Nuevo Simce de Inglés medirá comprensión oral y escrita del idioma,” 2010). The purpose of this test was to diagnose students’ level of English in order to orientate strategies aimed at the improvement of their learning (MINEDUC, 2011). However, it was later admitted that since the test is taken by students who are one year from finishing secondary school, it was difficult to do remedial work with them and the test should be taken by the end of primary school (MINEDUC, 2014).

The test implemented in 2010 corresponded to the TOEIC Bridge model designed and administered by ETS. Later applications in 2012 and 2014 were done on the basis of an adaptation of the Cambridge Key English Test (KET). In all the versions of the test, spoken and written production was left aside. Nevertheless, it was announced that the four language skills would eventually be assessed in an undetermined future application (MINEDUC, 2014)

For the 2010 English SIMCE Test, students had to score 64 points out of 100 in the listening comprehension part and 70 out of 100 in the reading comprehension part, in order to certify their A2 level according to the Common European Framework of Reference (CEFR). A2 was the minimum set by the Chilean Ministry of Education for students finishing their primary school. According to the CEFR standards, A2 learners can “understand phrases and the highest frequency vocabulary related to areas of most immediate personal relevance (...) [and] catch the main point in short, clear, simple messages and announcements,” regarding listening comprehension. In terms of reading comprehension said learners can “read very short, simple texts (...) [,] find specific predictable information in simple everyday material such as advertisements, prospectus, menus and timetables and (...) understand short simple personal letters” (Council of Europe, 2001).

On the other hand, for the 2012 and 2014 implementations of the test, candidates needed to score 70 points or above out of 100 in the test in order to obtain the A2 (70 to 89) or B1 (90 to 100) certification. Contrary to the 2010 test, no information was provided as to how high learners needed to score in the skills individually (MINEDUC, 2013 and 2015).

Taking this data into account, the national results have evidenced an increase over the years, yet this increase seems not to be significant enough considering that students take the examination three years after they should have attained the required level. In 2010, 11% of candidates reached A2 (no information was provided regarding the number of students reaching B1). Later, in 2012, 18% of examinees were certified (8.2% B1 and 9.6% A2). Finally, in 2014, almost 25% of students were certified (12.6% B1 and 12% A2) (MINEDUC, 2011, 2013 and 2015). This study covers the 2017 process. During this application, the test was taken by a limited number of schools chosen at random. At the moment of the writing of this article, the methods

or criteria used to select the schools had not been disclosed. The researchers tried to interview government authorities to discuss these details to no avail.

### *SIMCE and Washback*

Although not particularly in the case of the English variant, the washback the SIMCE tests have on teaching has already been reported on (Eyzaguirre & Fontaine, 1999). However, the studied impact has been largely connected to decisions made after the results of previous measurements have been made available to teachers and school authorities not to actions taken in order to obtain such results. In general terms, the published analysis of SIMCE's washback (although such terms have not been explicitly used) seems to indicate the aforementioned set of tests has had a mainly negative effect despite Ministry and constructors' original intentions. Among the negative effects caused by the SIMCE tests, researchers have identified stigmatization of students, teachers and authorities belonging to schools with poor results, exclusion of low achieving students who may decrease the general results, poor interpretation of results by teachers, school authorities and parents, exaggerated pedagogical attention on groups of students who are to be tested at the expense of the others, and the prioritization of the preparation for the next SIMCE rather than working on the weaknesses of the students who took the previous one (Ortiz, 2010). Additionally, teachers do not usually use the obtained data to revise their objectives or pedagogical actions (Flórez, 2015).

## METHODOLOGY

### *Design*

This study took a quantitative approach and had an exploratory nature. Data was obtained from a Likert scale questionnaire and processed using the SPSS software. Through the use of the aforementioned program, average scores and standard deviation were calculated for the inquiries related to each of the specific objectives.

### *Objectives*

The present study was designed considering the following objectives:

General objective:

To determine how the implementation of the 2017 English SIMCE test influences pedagogical actions taken by 11<sup>th</sup> year teachers of English in Puerto Montt, Chile.

Specific objectives:

1. To determine teachers' claim of understanding the purpose of the 2017 English SIMCE test.
2. To establish the 2017 English SIMCE test washback on lesson planning.

3. To describe the 2017 English SIMCE test washback on assessment.
4. To identify the 2017 English SIMCE test washback on decisions made by school authorities.
5. To identify teachers' perceptions regarding the use of the results of the 2017 English SIMCE test.

### *Delimitations*

This research was carried out in Puerto Montt, Chile. The researchers gathered the data between the months of June and August of 2017. For the purpose of this study, the researchers analyzed solely the participants' answers to the online questionnaire described below. Originally, classroom observation, analysis of teaching materials and interviews with school authorities were also considered. However, due to lack of funding, those actions were discarded. The aforementioned actions might be considered in future studies on the topic.

### *Participants*

Thirty-nine instructors of English as a Foreign Language, who were teaching English in 11<sup>th</sup> year during 2017 in the Chilean city of Puerto Montt, initially participated in the research. These teachers were selected as 11<sup>th</sup> grade would be the year assessed by the latest version of the English SIMCE test. The online questionnaire was originally sent to 53 instructors teaching English in the aforementioned year, on June 6<sup>th</sup>. These teachers belonged to 4 private schools, 8 public schools and 23 government subsidized private schools. Within 65 days, 39 teachers had answered the questionnaire. The remaining teachers neither answered the questionnaire nor did they confirm the reception of it.

Out of the 39 teachers who answered the questionnaire, 7 belonged to purely private schools; 10, to public schools; and, 22, to government subsidized private schools. For the purpose of this study, public and purely private school data was discarded as the number of teachers was not representative.

### *Instrument*

Participants answered an online questionnaire in the form of a Likert scale. The questionnaire contained 17 questions that were answered in a scale from *Completely disagree* to *Completely agree*. Questions were posed in accordance to studies mentioned in the literature review and intended to check the impact of the English SIMCE test on teachers' actions regarding planning and assessment. Additionally, teachers' perceptions in connection with the exam were also measured. Finally, more

global decisions, such as selection of contents or exclusion of low-achieving students, were also inquired. Scores to answers were assigned on a scale from 1 (*Completely disagree*) to 5 (*Completely agree*), 1 showing a low impact of the English SIMCE test on the given domain, while 5 showed a high impact.

During the analysis of answers, questions were grouped according to the specific objective(s) they were aimed at answering. Some inquiries related to more than one specific objective as the particular domain of teaching they were connected with could affect different aspects of teachers' practices.

### *Procedures*

The team of researchers was formed in October 2016. It consisted of two teachers of English and one teacher of Biology. Between November 2016 and April 2017, relevant literature was reviewed and the instrument was designed and validated. From April to June, researchers gathered contact information of the teachers who would answer the online questionnaire. The gathered data was processed and analyzed from November to January 2017. Finally, the present article was written during January and February 2017.

### *Limitations*

The present research has the following limitations detected by the researchers. First, the nature of the online questionnaire made it impossible to ensure the participation of all the instructors who would be teaching English in 11<sup>th</sup> year. Second, as the online questionnaire was the sole source of data, it may have shown results that deviate from the real washback that the SIMCE test may have. For future implementations, classroom observation and analysis of teaching materials could complement the data gathered through the questionnaire and increase its validity.

## RESULTS AND DISCUSSION

Teachers' answers were analyzed in terms of the specific objectives stated above. Specific objective 1: Determine teachers' claim of understanding the purpose of the 2017 English SIMCE test.

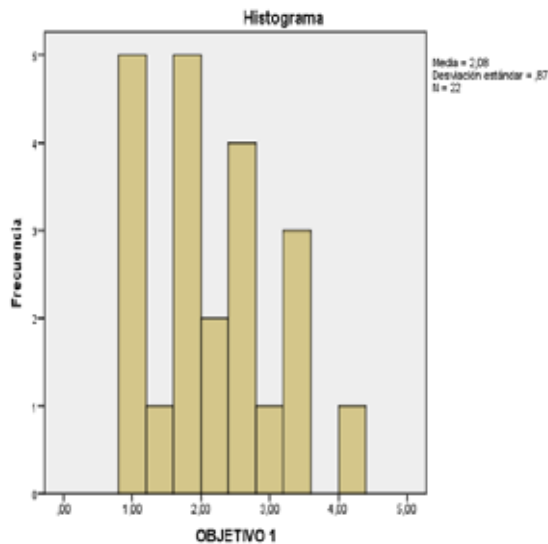
The present specific objective was answered by question 1 (see appendix). Respondents averaged 3.68 points. The most common answer was *Completely agree* with 7 respondents. Dispersion of answers showed a tendency towards understanding the purpose of the examination. The nature of the question does not allow concluding whether the teachers actually understood the purposes of the test, as it only considers their claim. In fact, the analysis of Specific objective 5 seems to indicate teachers were not completely familiar with the objectives underlying the implementation of the SIMCE test.

Specific objective 1	
Question 1	
Average	3.68
Standard deviation	1.25

Answer	Number of respondents
Completely disagree	2
Partially disagree	1
Neutral	4
Partially agree	4
Completely agree	7

Specific objective 2: Establish the 2017 English SIMCE test washback on lesson planning.

The second specific objective was connected with questions 6, 8, 10, 11 and 15 (see appendix). From the histogram, a tendency towards low-valued answers can be noticed. This tendency shows that the teachers included in the study, in general terms, seem not to attribute influence of the SIMCE test on their teaching methods, the contents to be taught or the priority of language skills in their planning. Additionally, the majority state that preparing their students for the high-stakes examination does not add to their workload. Unfortunately, it is not possible to determine whether this means that they did not give extra preparation to their students or that they did, but it was not significant within their regular workload.

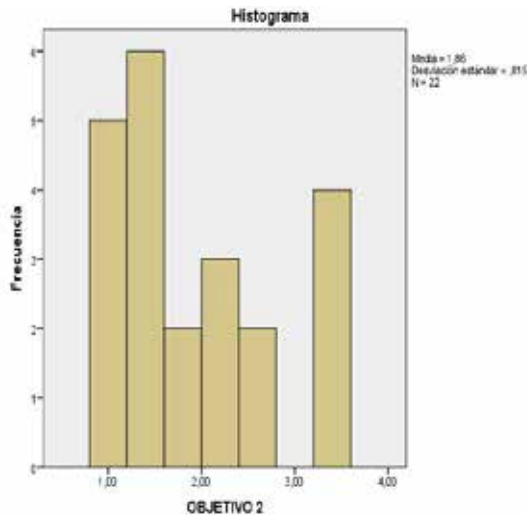




Specific objective 2	
Questions 6, 8, 10, 11, 15.	
Average	2.08
Standard deviation	.87

Specific objective 3: Describe the 2017 English SIMCE test washback on assessment.

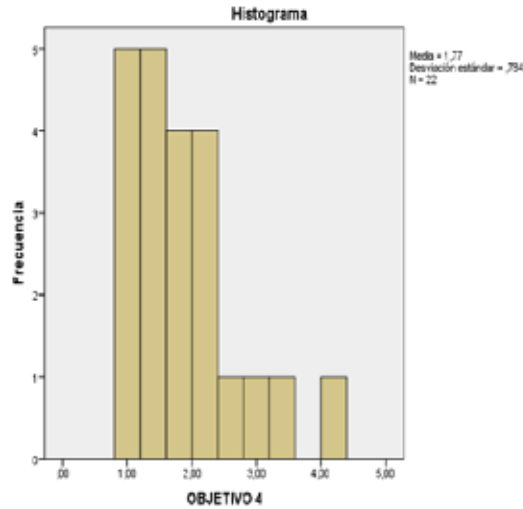
The third specific objective was connected with questions 7, 10, 12, 13 and 14 (see appendix). Similar to the previous specific objective, answers seem to trend towards low-valued answers. These answers indicate that respondents did not feel their assessment methods were influenced by the English SIMCE, nor did they have to use assessment instruments they did not agree with or that their students were not familiar with.



Specific objective 3	
Questions 7, 10, 12, 13, 14.	
Average	1.86
Standard deviation	.81

Specific objective 4: Identify the 2017 English SIMCE test washback on decisions made by school authorities.

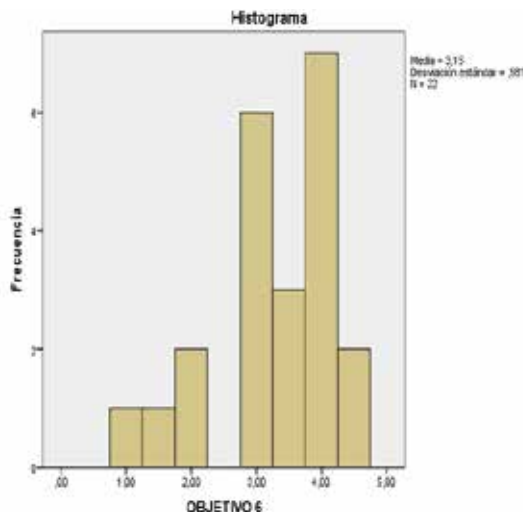
The fourth specific objective was connected with questions 8, 9, 10, 16, and 17 (see appendix). In this case, there is again a marked tendency towards low-valued answers. Most respondents stated that the implementation of the 2017 English SIMCE did not have an impact on decisions made upon their practices by school authorities. Such decisions were itemized into the contents to be taught, excluding low-achieving students from the process, addition to the regular workload, distribution of teaching materials designed to cope with the exam and suggestions in terms of how to teach 11<sup>th</sup> year students.



Specific objective 4	
Questions 8, 9, 10, 16, 17.	
Average	1.77
Standard deviation	.78

Specific objective 5: Identify teachers’ perceptions regarding the use of the results of the 2017 English SIMCE test.

The fifth specific objective was connected with questions 2, 3, 4 and 5 (see appendix). Opposed to all previous specific objectives, respondents’ answers showed a tendency towards high-valued answers. In the case of the last specific objective, teachers mostly state that the results will show how much their learners know, but at the same time those results will be used to judge their teaching practices by school authorities, parents and students.



Specific objective 5	
Questions 2, 3, 4, 5.	
Average	3.14
Standard deviation	.98

## CONCLUSIONS AND IMPLICATIONS

In general terms, this application of the SIMCE test seems to be harmless in terms of the negative washback it could have created on the surveyed teachers. When isolating the components of their teaching practices that could have been affected according to previous research, teachers' answers tended to claim that the test did not have an impact on them. As previously stated, the 2017 version of the English SIMCE test was applied to a limited number of schools that were chosen at random and whose teachers were not told about the application in advance. Thus, it is feasible to think that such conditions may partially be the reason for such a low perceived impact, i.e. not being aware of the upcoming testing diminishes the impact the application of the test will have on teachers' practices. Further research is needed to prove whether different conditions may lead to different results in upcoming versions of the examination.

## REFERENCES

- ALDERSON, J. C., & WALL, D. 1993. "Does washback exist?" *Applied Linguistics*, 14: 115-129.
- BROWN, H. D. 2004. *Language assessment: Principles and classroom practices*. White Plains, NY: Pearson Education.
- CHENG, L., WATANABE, Y. & CURTIS, A. 2008. *Washback in language testing: Research contexts and methods*. Mahwah, NJ: Lawrence Erlbaum Associates.
- CHENG, L. 2008. "The Washback Effect of a Public Examination Change on Teachers' Perceptions Toward Their Classroom Teaching". In L. CHENG, Y. WATANABE, & A. CURTIS (Eds.), *Washback in language testing: Research contexts and methods*. Pp. 3-17. Mahwah, NJ: Lawrence Erlbaum Associates.
- COUNCIL OF EUROPE. 2001. *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge, U.K: Press Syndicate of the University of Cambridge.
- EYZAGUIRRE, B. & FONTAINE, L. 1999. "¿Qué mide realmente el SIMCE?" *Estudios Públicos*, 75: 107-161.
- FISH, J. 1988. *Responses to mandated standardised testing*. Unpublished doctoral dissertation, University of California, Los Angeles.
- FLÓREZ, M. 2015. "Validity and equity in educational measurement: The case of SIMCE". *Psicoperspectivas. Individuo y Sociedad*. 14 (3): 31-44.
- MINEDUC. 2011. *Resultados de Inglés para Docentes y Directivos*. Ministerio de Educación. Gobierno de Chile.

- MINEDUC. 2013. *Resultados de Inglés para Docentes y Directivos*. Ministerio de Educación. Gobierno de Chile.
- MINEDUC. 2014. *Estrategia Nacional de Inglés 2014.2030*. Ministerio de Educación. Gobierno de Chile.
- MINEDUC. 2015. *Resultados de Inglés para Docentes y Directivos*. Ministerio de Educación. Gobierno de Chile.
- MUÑOZ, R. 2017. "The effect of washback on EFL learners' attitude towards tests". *Studies in English Language Teaching*. 5 (3): 516-530. DOI: <http://dx.doi.org/10.22158/selt.v5n3p516>
- NOBLE, A. J., & SMITH, M. L. 1994. *Measurement-driven reform: Research on policy, practice, repercussion* (Tech. Rep. 381). Tempe, AZ: Arizona State University, Center for the Study of Evaluation.
- EMOL. 2010. Nuevo Simce de Inglés medirá comprensión oral y escrita del idioma. [online]. Available at <http://www.emol.com/noticias/nacional/2010/07/23/426601/nuevo-simce-de-ingles-medira-comprension-oral-y-escrita-del-idioma.html> [Consulted on 08/06/2018].
- ORTIZ, I. 2010. *25 años del SIMCE*. Cide. Centro de Investigación y Desarrollo de la Educación. Universidad Alberto Hurtado.
- PAN, Y. 2009. "A review of washback and its pedagogical implications". *VNU Journal of Science, Foreign Languages* 25: 257-263.
- QI, L. 2008. "Has a High-stakes Test Produced the Intended Changes?". In L. CHENG, Y. WATANABE, & A. CURTIS (Eds.), *Washback in language testing: Research contexts and methods*. 171-190. Mahwah, NJ: Lawrence Erlbaum Associates.