

## ASSESSMENT IN THE ENGLISH LANGUAGE CLASSROOM IN CHILE: EXPLORING THE WASHBACK EFFECT OF TRADITIONAL TESTING AND ALTERNATIVE ASSESSMENT ON SEVENTH GRADE STUDENTS

LA EVALUACIÓN DEL INGLÉS EN LA SALA DE CLASES EN CHILE: UNA  
EXPLORACIÓN DEL EFECTO *WASHBACK* DE PRUEBAS TRADICIONALES  
Y EVALUACIONES ALTERNATIVAS EN ESTUDIANTES DE 7° GRADO

PAOLA MUÑOZ\*, MAURICIO VÉLIZ-CAMPOS\*\*,  
LEONARDO VÉLIZ\*\*\*

### Abstract

Assessment has long been regarded as an integral part of the learning process. In the field of assessment, washback effect is understood as the impact of tests (or any assessment procedure) on the learner, the learning process, teachers, or such like. Thus, the purpose of the present study was to determine the washback effect of two different types of assessment procedures, namely, a traditional test and an alternative assessment procedure (a project) in an intact sample of 32 seventh-grade students, from a subsidised school in Chile. Through a mix-methods approach, quantitative data were gathered through a self-reported-washback survey, administered after a traditional test was given to the participants and after an English language project was carried out; qualitative data regarding the perceived effects caused by the test and the language project were gathered through a focus group interview. A series of *t*-tests was performed for quantitative data gathered through the washback surveys, while content analysis was used for the qualitative data emerging from the focus group interview. The results suggest that both procedures are positively valued, with the alternative assessment pro-

\* Licenciada en Educación y Profesora de Inglés. Magíster en TESOL, paolabeatrizmu@gmail.com

\*\* Doctor en Educación. Profesor Asistente en la Facultad de Educación de la Universidad de Talca, mauricio.veliz@utalca.cl.

\*\*\* Head of School of Education Excelsia College, Australia, leonardo.veliz@excelsia.edu.au

cedure being held in higher regard as far as motivation, anxiety and strategy use are concerned, as became evident in the focus group interview.

**Keywords:** Alternative assessment, anxiety, motivation, strategy, traditional testing procedures.

## Resumen

La evaluación ha sido concebida como aspecto fundamental de los procesos de aprendizaje. En este ámbito el efecto denominado *washback* se entiende como el impacto de las pruebas (o de cualquier procedimiento de evaluación) en los aprendientes, el proceso de aprendizaje, los profesores u otros. Es así como el propósito de la presente investigación fue determinar el efecto de dos procedimientos evaluativos: una prueba tradicional y un procedimiento de evaluación alternativo (un proyecto), en una muestra intacta de 32 estudiantes de séptimo grado de una escuela subvencionada de Chile. Mediante un método mixto, se emplearon datos cuantitativos recogidos a través de una encuesta de autorreporte de efecto *washback*, administrada después de la prueba tradicional y de culminar el proyecto; asimismo, se recogieron datos cualitativos en un grupo focal vinculados a los efectos percibidos por los participantes después de la implementación de ambos procedimientos. Se aplicó una serie de pruebas *t* para los datos cuantitativos, mientras que, para los datos cualitativos, se empleó análisis de contenido. Los resultados indican que ambos procedimientos evaluativos son percibidos positivamente, aunque se observa una mejor valoración asociada al procedimiento de evaluación alternativo, particularmente en lo que se refiere a dimensiones de motivación, ansiedad y estrategia, lo que se evidenció especialmente en grupo focal.

**Palabras clave:** Evaluación alternativa, motivación, estrategia, procedimiento de evaluación tradicional.

## Introduction

AS EARLY AS IN the second century, oral evaluations were performed with the purpose of selecting the most suitable candidates to work as civil servants in China (Alcaraz, 2015). Thus, assessing human abilities and assessment *of or for* learning in education has a long-standing tradition, a process that, for the most part, informs subsequent decision making. English language learning is by no

means an exception in that assessment involves collecting information about language learners' learning experiences to later introduce the necessary adjustments to language teaching and/or the language planning structure (Brown, 2004).

Broadly speaking, two major types of assessment procedures can be distinguished, namely traditional and alternative assessment. The former, which has a much longer history, is generally characterised by testing practices through which objective and measurable data are gathered about some aspect of reality (Fulcher, 2010). The latter, on the other hand, has been particularly focused on a set of practices aimed at measuring reality in less objective ways (Fulcher, 2010). These two different assessment procedures, however, must not be conceived of as two opposing ends of a continuum; but rather, two complementary measures of language ability. Alternative assessments have also gained greater recognition in education at large and in second language learning in particular.

In the field of assessment, washback effect is, according to Akpınar and Cakildere (2013), understood as the impact of tests (or any assessment procedure) on the learner, the learning process, or teachers. The authors claim that washback has become a popular area of study within educational research because of the powerful implications for pedagogy. As Cheng and Curtis (2012) state, washback is also described as the consequences of testing and evaluation used in the classroom; Akpınar and Cakildere (2013) suggest that washback is often used as a synonym for 'impact' on, 'effect' on or 'consequences' in educational contexts, which can be positive or negative, as suggested by Alderson and Wall (1993). Thus, not only does testing - and assessment at large - collect information about the students' learning and the learning experience, but also it invariably causes effects on the learners on the learning experience, on how the curriculum is enacted, and potentially, on the learning environment. Tayeb, Aziz, Ismail, and Khan (2014) studied the washback effects of the General Secondary English Examination (GSEE) on teaching and learning of high-stakes test for secondary school students in Yemen. The authors concluded that clear evidence was found in support of the washback effect of the exam on various aspects of the language teaching-learning process, particularly

on what and how teachers teach, and on what and how the learners learn. This clearly points to the fact that traditional testing practices tend to cause, more often than not, high levels of anxiety in learners and in teachers alike. This undeniable fact has raised numerous questions about the merits of, for instance, high-stakes tests, being employed across the world, and their impact on students and teachers.

Alternative assessment procedures, however, are viewed positively on account of their ability to provide learners with an opportunity to demonstrate what they know and, most importantly, what they can do (Al-Ruqeishi & Al-Humaidi, 2016); also, they are more authentic in their elicitation of meaningful communication and produce more meaningful feedback to students (Brown, 2004). Likewise, Petre (2017) regards alternative assessment as innovative and motivating. As the author puts it, “learners need authentic, dynamic, innovative and creative tasks which extend students’ perspectives and increase motivation for learning” (p. 159). Consequently, alternative assessment procedures are now being favoured on account of various grounds, among which is its potential to produce reduced anxiety levels (washback effect) in learners; its capacity to assess actual performance and ensure face and content validity; its meaningfulness and authenticity, which explains, as Inbar-Lourie and Donitsa-Schmitt (2009) claim - at least in part - why increased efforts are made to foster the use of alternative assessment in large scale educational contexts, as is the case of Israel.

Safa and Goodarzi (2014) conducted a study in Iran on the washback effects of task-based language assessment (TBLA) on EFL learners’ grammar development. A control and an experiment group received the same grammar instruction over ten sessions; the experimental group took a task-based grammar quiz every three sessions, while the control group took traditional grammar quizzes. The results revealed that TBLA had a positive washback effect on the learners’ subsequent grammar learning process.

Drawing on a mixed-methods methodology, this study used a washback survey along with a focus group interview to achieve a twofold purpose: First, it sought to determine the washback effect of administering both alternative assessment procedures and tradi-

tional tests in an intact seventh grade class from a subsidized school in Chile; second, it aimed at gathering students' perceptions of the effects of both types of assessment on (their) motivation, anxiety, and strategy use. Thus, the research questions this study aimed to answer are: (i) What are the washback effects of alternative and traditional assessment procedures administered to seventh grade students from a subsidized school in Chile? and (ii) What are seventh graders' views on the the washback effect produced by the administration of alternative assessment procedures and of traditional tests?

Given that much of the washback-related research conducted thus far has focused primarily on high stakes tests (Shohamy, 2017), the political uses (and effects) of language testing (Shahomy, 2001), the effects of test preparation, usually at college/university level (Ma & Cheng, 2016), with no reference to Chilean (school) contexts, this study provides an important opportunity to advance the understanding of the possible washback effects of two different assessment procedures on school children in an English as a foreign language classroom in Chile, which corresponds to the the most widely used scenario for English language learning.

## **Review of literature**

### *Second language assessment*

Assessment plays a fundamental role in education at large and in foreign language learning contexts in particular (Alderson, Brunfaut, & Harding, 2017) as it involves gathering information about what learners know and can do before, during, or after a learning cycle, with a view to making decisions accordingly. Assessment involves testing, measuring, collecting and combining information, and providing feedback (Shohamy, 2017). In the realm of foreign language learning, assessment entails eliciting samples of L2 performance from a learner under certain conditions, to later 'pass a judgment' and make subsequent decisions about the learning process and those who take part in it (Purpura, 2016).

While formative assessment most often takes place inside the classroom since it focuses on the actual learning process, summative assessment tends to take place at the end of a learning cycle and is aimed at measuring learners' learning outcomes. State-driven high-stakes tests in foreign language learning contexts serve the purpose of gathering a good deal of data and, at the same time, implementing accountability mechanisms. As Meckes and Carrasco (2010) state, the national standardised high-stakes English language test, administered every two years nationwide in Chile, intends to improve the quality and equity of education by "providing data about learning outcomes at the national and school levels in order to monitor and inform decision-making, provide feedback to schools, improve teaching practices, and foster accountability, parental, and school community involvement" (p. 234). In so doing, high-stakes - and classroom summative - foreign language tests also have the potential to shape teachers' teaching methodologies, alter foreign language syllabi, and place pressure on teachers, administrators, students, and parents (Williams, 2014), although some positive effects produced by tests have also been reported (Tzagari & Cheng, 2017).

At classroom level, traditional tests - understood as paper-and-pencil tests (McNamara, 2000), high-stakes, discrete-point, multiple choice tests (Fox, 2017) - have gained increased power; so much so that at times language teaching is neatly structured around past papers or practice tests, (Williams, 2014), which explains why traditional tests are mistrusted. As Hughes (2003) argues, "teaching is, after all, the primary activity; if testing comes in conflict with it, then it is testing that should go, especially when it has been admitted that so much testing provides inaccurate information" (p. 4). Hughes (*ibid.*) moves on to claim that students' true abilities are not always reflected in the test scores they obtain; by the same token, he contends that language abilities are not easy to measure as, for instance, physical sciences are, which explains in part why alternative assessment procedures have become pervasive over the last two decades (Fox, 2017), at both macro- and micro-levels of language education.

*Alternative assessment*

As has been hinted earlier, traditional assessment has been bitterly criticized because of its unvaried forms of assessing students' performance, which do not satisfactorily consider the individual skills learners have. Brown (2004) described alternative assessment procedures as more authentic methods to assess students because of their elicitation of meaningful communication in contrast to what traditional assessment offers. Some of the most common alternative assessment procedures (or alternatives *in* assessment) are "criterion-referenced checklists, reading response journals, learning logs, poster presentations" (Fox, 2017, p. 135), together with portfolios, projects, among others. As the name suggests, alternative assessment procedures and techniques provide alternative forms of organisation of the assessment activities (Petre, 2017) and of viewing the language learner, and the learning process. When implementing alternative assessment, learners are part of the process that measures their learning, which involves the application of knowledge and skills in real life contexts. This type of assessment procedure, as Al-Ruqeishi and Al-Humaidi (2016) contend, demands students' creativity and effective performance.

Petre (2017) conducted a study on alternative assessment procedures where the author analysed students' possible preferences for (i) alternative assessment strategies, alternative assessment methods, traditional assessment strategies, traditional assessment methods; and (ii) the opportunities for the practical application of the skills and knowledge learned through assessment using alternative methods and strategies or traditional methods and strategies. The findings indicated that students preferred assessment procedures such as projects and portfolios, in that order, since the former provided the students with ample opportunities for actually applying the skills developed and the knowledge acquired. Similarly, portfolios have been increasingly used, in different forms and for different purposes, and their benefits have been widely promoted (Fox, 2017), one of which has been their potential for test-related anxiety reduction, as Contreras, Véliz-Campos and Veliz, argue (2019), amongst other potential benefits.

Portfolios, together with projects, have proven particularly helpful in enhancing learners' language development and assessing their learning. For instance, Ghoorchaeei, Tavakoli, & Ansari (2010) investigated the effect of portfolio assessment on Iranian EFL students' writing abilities and other sub-skills. The findings suggested that portfolio assessment positively affected students' overall writing achievement, in conjunction with their writing focus, development, and organization skills, aside from enhanced vocabulary development. These findings corroborated Aydin's (2010) study, which explored 204 EFL Turkish students' perceptions on portfolio assessment, where the results suggested that portfolios contributed to their vocabulary and grammar knowledge, reading, research, and writing skills. Similarly, portfolios and projects have proven useful tools for feedback provision, which is why some learners express their strong preferences for this type of alternative assessment as it successfully mitigates their test-anxiety level (Lam, 2014), however tiring, long, and time-consuming it may be.

Notwithstanding the tendencies of both traditional testing and alternative assessment to produce more or less of an impact on language learners, it must be pointed out that all forms of assessment have potential shortcomings and that all forms of assessment, particularly summative traditional tests and high stakes tests, tend to have a greater effect, at least, on the test-taker (Hughes, 2003). This effect, termed *washback effect* (Safa & Goodarzi, 2014; Zhan & Andrews, 2014), can thus be positive or negative.

### *Washback effect*

Washback can be loosely understood as “the influence of testing on teaching and learning” (Xie & Andrews, 2013). This definition is a simplified, yet useful version of what other authors had previously posited in terms of washback being the impact of testing on both local and broader learning contexts. Even though *impact* and *effect* (or *influence*, for that matter) may be used interchangeably in reference to washback, MacNamara (2000 as cited in Cheng, Sun & Ma, 2015), argues that *impact* makes reference to the macro-level effect of tests, e.g. that



of national testing policies, whereas *washback* refers to the micro-level effect, i.e. at classroom level. Recent research has paid particular attention to washback in classroom-based contexts. Indeed, Cheng, Sun and Ma (2015) claimed that washback deals with the effect of the “introduction of a new test on classroom teaching and learning”.

However straightforward the term washback may seem, the literature shows a degree of opacity as to how exactly testing influences teaching and learning (Xie & Andrews, 2013) as there are a number of manifestations of washback both in micro and macro contexts, together with the wide array of perceptions of how students perceive washback (Cheng & DeLuca, 2011; Xie & Andrews, 2013). Broadly speaking, the literature suggests that washback can be either positive or negative (Cheng, Su & Ma, 2015). Positive washback may, for instance, manifest itself in teachers and learners working collaboratively towards learning objectives as a result of the administration of a given testing/assessment procedure (Cheng, Sun & Ma, 2015). Negative washback, on the other hand, occurs when results are not correctly interpreted or are simply misused (Cheng, Sun & Ma, 2015). Therefore, teachers must aim at enhancing the positive side of washback and reducing the negative side of it (Cheng, Sun & Ma, 2015).

Recent research studies have examined the influence of testing, especially in contexts where high-stakes tests are pervasive at school level in different countries such as Australia, China, Canada, and Iran (Collins, Reins & Stobart, 2010; Polesel, Rice & Dulfer, 2014). These studies have analysed the impact of high-stakes testing on the curriculum as well as on classroom practices. Polesel, Rice and Dulfer (2014) conducted a study aimed at exploring the impact of the National Assessment Program – Literacy and Numeracy (NAPLAN), a high-stakes test administered in Australian schools. More than 8,000 educators were asked about the effects of the implementation of NAPLAN. The findings indicated that much of the curriculum coverage and classroom practices in Australia are largely test-driven. Indeed, more than 75% of teachers felt that their teaching was devoted to test preparation. Similarly, Collins, Reiss and Stobart (2010) explored the opinions of Year-6 science teachers in England and Wales regarding the use of standardized tests and the abolition of national testing

in Wales. While the participants reported a positive attitude towards high-stakes testing since it helped teachers to narrow down the curriculum, the teachers also underscored that national test scores “failed to provide accurate information about pupil’s achievements and attainment” (p. 283), and that not having standardized testing would change their teaching methodology, thus enhancing students’ motivation and interest in learning science.

Washback also has implications for test validity and test fairness (Cheng, Sun & Ma, 2015; Collins, Reiss & Stobart, 2010). Cheng and DeLuca (2011) conducted a study in which they explored the opinions of test takers on the factors that influence test validity. The 59 participants highlighted the influence of testing on a) test administration and testing conditions, b) timing, c) test structure and content, d) scoring effects, e) preparation and test-taking strategies, f) test purpose, g) psychological factors. In their study, the participants “mentioned issues related to preparation and test-taking strategies most frequently, suggesting their belief in coaching and test preparation” (Cheng & DeLuca, 2011, p. 116). Moreover, the participants expressed that test washback’s unpredictability may be linked to teachers’ lack of understanding of learners’ expectations and beliefs in a testing situation. Hence, the evidence seemed to suggest that preparing students for tests and knowing students’ expectations added to test validity.

Overall, the exploration of the washback effect has been done mostly for high-stakes and large-scale testing. Research studies have suggested that tests produce learning directed to the test, which does not necessarily mean negative consequences. Moreover, as several authors contend (Cheng & DeLuca, 2011; Polesel, Rice & Dulfer, 2014), there are a number of factors that may be considered when observing the effect of tests in learners in both cognitive and emotional aspects.

### **Instruments of data collection**

The research methodology that informs the present study is grounded in the principles of what has come to be known as ‘the third research

method' (Creswell, 2009) or mixed-methods approach. The methodological justification for combining two research methods that are, by and large, seemingly distinct lies at the heart of the benefits gleaned from using quantitative and qualitative methods (Creswell, 2009).

Quantitative data relating to the participants' perceived washback effect of both an alternative assessment procedure and a traditional testing procedure were collected using a washback survey. Qualitative data were gathered through a focus group interview. The survey was administered in Spanish in order to avoid any possible linguistic misunderstanding. Similarly, the focus group interview was held in Spanish, thus allowing for greater depth and an honest voice for the participants.

The washback survey comprised two sections: The first probed into the participants' relationship with the English language and their learning process; the second section consisted of 20 items on a four-point Likert scale. Because washback is often equated with "the influence of testing on teaching and learning" (Xie & Andrews, 2013), which can be positive or negative (Shohamy, Or & May, 2017), three washback-related categories were imbedded in the survey, viz. motivation (items 1-7), anxiety (items 8-15), and strategy use potential (items 16-20) (see Appendix 1). Figure 1 below presents a sample of the washback survey.

II. Marque con una "X" su respuesta.

**MA:** Muy de acuerdo  
**A:** De acuerdo  
**D:** En desacuerdo  
**MD:** Muy en desacuerdo

1	Motivación	MA 1	A 2	D 3	MD 4
1.1	Me sentí motivado a trabajar en el tema/contenido de esta evaluación.				
1.2	Disfruté estudiando/trabajando para esta evaluación.				

Figure 1. Sample of washback survey. Source: Authors' table.

The washback survey was validated through various means, one of which was expert judgement (Meyer & Booker, 2001). First, it was later sent to an external judge with a validation instrument intended to rate the survey's clarity, relevance, and coherence. Minor changes were subsequently introduced as a result as the scores given to the survey items, which were for the most part considerably high for all three evaluative criteria. Minor changes entailed the wording of a couple of items and the need to rearrange the survey items in the second application in order to avoid a possible sense of repetition in the participants.

Credibility was largely based on triangulation, allowing for the analysis of data from the washback survey, administered twice, and the focus group interview, which was held two days after the second survey had been administered. The focus group interview was audio recorded and later transcribed *verbatim* using pseudonyms for the participants, and colour coded for the different emerging themes tapping into the crux of the research questions.

Lastly, the assumptions have been hinted above and are described as follows: (i) assessment procedures have an impact on learners, teachers, the learning process, either separately or in a combined form; (ii) the impact it has on the various participants and processes can be positive or negative; and (iii) the extent to which assessment procedures impact the various participants and processes may vary depending on the nature of the assessment procedure.

### *The procedure*

The washback survey was administered twice; first (Time 1), it was administered once the participants had carried out the alternative assessment procedure - the project -, after the course content had already been covered, while for the second time (Time 2), it was completed once the participants had taken the traditional test, preceded by the teaching of the relevant content. On both occasions, the researcher carefully walked the participants through the survey for any necessary clarification. All 20 items were shuffled around in the Time 2 survey in order to guarantee the participants' attention to the items, and thus, avoid reliance on their previous responses in Time 1 survey. In each

survey administration instance, the researcher asked the participants to complete the survey thinking of the assessment procedure they had just taken or completed. Two days after the Time 2 survey had been administered, the focus group interview was conducted. Figure 2 illustrates the entire procedure.

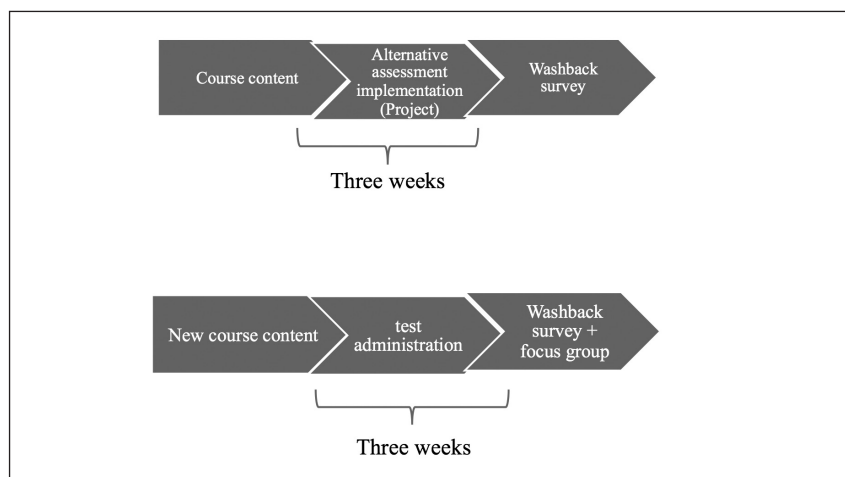


Figure 2. The procedure. Source: Authors' figure

The alternative assessment procedure consisted in a small-scale project wherein the participants selected a movie or a book of their preference and wrote reviews for it over a three-week period, continuously assisted by their teacher and applying the course contents covered during that time-period. The test, on the other hand, comprised reading, writing, and listening, with multiple choice questions and five open-ended questions.

Prior to commencing the data-gathering process, school permission to carry out the investigation was sought and granted formally. Similarly, informed consent in Spanish from the participants' parents was sought in writing, which clearly described the purpose of this study, the importance of the students' participation, and assured confidentiality and freedom to withdraw their participation in the study at any time.

### *The participants*

The study was conducted at a subsidised non-profit faith school located in San Pedro de la Paz, Concepción-Chile. Although the school is not bilingual, English language teaching is enthusiastically promoted. The study used a non-probability convenience sample of 40 seventh grade students (21 females and 19 males), six (three males and three females) of whom took part in the focus group interview on a voluntary basis. For data analysis purposes, eight cases were not considered as incomplete data had been provided. The sample was made up of an intact class one of the researchers taught at the time. The participants' age ranged from 13 to 14.

## **Results and discussion**

### *Measures*

As far as the the washback effect scale is concerned, the instrument proved highly reliable at time 1 ( $\alpha = .87$ ) and time 2 ( $\alpha = .90$ ). Similarly, virtually all the dimensions were also highly reliable, namely (i) motivation ( $\alpha_{t1} = .79$ ,  $\alpha_{t2} = .90$ ), anxiety ( $\alpha_{t1} = .75$ ,  $\alpha_{t2} = .81$ ), and strategy use potential ( $\alpha_{t2} = .86$ ). The only exception was the last factor at time 1, which was only barely reliable ( $\alpha_{t1} = .60$ ). The mean was computed for each factor and the total scale at both time 1 and time 2, for both the descriptive statistics and analyses. Thus, the higher the scores, the higher the level of the latent factor measured, the only exception being anxiety, wherein higher values indicate lower anxiety levels.

### *Research question 1: Washback effects of alternative and traditional assessment procedures administered to seventh graders*

We ran a series of *t*-tests to analyze the differences in the measures at each time and to compare the mean scores between times. In the specific case of the comparisons at both time 1 and time 2, we use the Bonferroni correction for multiple comparisons (Armstrong, 2014).

Therefore, given that there were three non-independent comparisons, the  $p$ -value was set at  $\alpha = .05/3$ . In other words, to reject the null hypothesis of equality between means, the  $p$ -value would be lower than .017 instead of .05 in order to decrease the probability of Type I error derived from this type of comparisons. All the analyses were conducted using the software Stata v.13

Descriptive statistics are presented in Table 1. As can be seen, when comparing the scores for Time 1, no significant differences among the factors can be observed. Specifically, motivation was not different from anxiety,  $t(27) = -.11, p = .917$ , or strategy use potential,  $t(27) = -1.39, p = .177$ ; and anxiety was not statistically different from strategy use potential,  $t(27) = -1.19, p = .243$ . At time 2, the same pattern of results can be observed: Motivation was not significantly different from anxiety,  $t(27) = -.39, p = .697$ , or strategy use potential,  $t(27) = -.33, p = .746$ ; and anxiety was not significantly different from strategy use potential,  $t(27) = .12, p = .906$ .

When comparing the scores between both times, no significant differences in any of the factors or in the total score can be observed. In particular, the mean was not significantly different at time 1 and time 2 for motivation,  $t(27) = 1.16, p = .256$ , anxiety,  $t(27) = .55, p = .584$ , strategy use potential,  $t(27) = 1.59, p = .123$ , or the total score,  $t(27) = 1.34, p = .192$ .

**Table 1.** Descriptive statistics.

	M	SD
Motivation (t1)	3.07	0.56
Anxiety (t1)	3.08	0.56
Strategy use potential (t1)	3.21	0.51
Washback Effect Scale (t1)	3.11	0.47
Motivation (t2)	2.95	0.66
Anxiety (t2)	3.00	0.64
Strategy use potential (t2)	2.99	0.72
Washback Effect Scale (t2)	2.98	0.54

Source: authors' table.

The quantitative findings emerging from the analysis of the differences in measures at Times 1 and 2 reveal some interesting aspects of the effects of traditional testing and alternative forms of assessment on students' anxiety, motivation and strategy use. Firstly, it was observed that there was no significant statistical difference in the washback effects of testing and alternative assessment on students at the two measured instances. Although this may be easily interpreted as an instance where there was no quantitative change in behaviour (Shawn & McMillion, 2008), it is important to bear in mind that what influences changes in learning patterns and learning behaviour can be anything from a wide variety of factors pertaining to the individual, or the learner, and to the wider social context in which they are embedded (Zhao, 2015). This means that washback effects of formal testing or alternative assessments are not to be attributed only to features and qualities of tests, but to a range of conditions that may facilitate or hinder test performance. Thus, it is assumed in this study that the lack of quantitative changes in students' motivation, anxiety and strategy use in both instances has to be understood within a specific context by paying particular attention to what, apart from the actual differences in measuring students' learning, might have contributed to a situation where no washback effects were perceived as a result of a test and an alternative form of assessment.

It could be speculated that one factor which may have influenced these quantitative findings was the relationship between the teacher-researcher and the students. The rapport that a teacher develops with students is important, and often necessary, to create appropriate learning conditions and to enhance students' overall language learning experience (Hussain, Nawaz, Nasir, Kiani & Hussain, 2013). This may be a potential factor that influenced the perceived washback effects of testing and alternative assessment on students' levels of anxiety, motivation and uses of strategies. What is important to highlight is, irrespective of whether or not the teacher-student relationship was an influencing factor, that although the use of alternative forms of assessment did not seem to have any quantitative effects on the above measures, students did report significant qualitative changes in their motivation, anxiety and strategy use.



*Research question 2: Students' views on the implementation of alternative and traditional assessment procedures*

The participants seemed to concur on the perceived positive effects produced by the use of alternative assessment procedures as far as motivation, anxiety, and strategy use, despite the virtually non-existent statistical difference, as reported above. Indeed, when asked about whether the use of the project as an assessment procedure made them feel motivated or not, Peter remarked that “...*in the projects..., at least in my case, it is something I really like, so in general terms I learn easily when I do something I like, I feel more motivated to learn it*” [free translation]. In a similar vein, Alice added that projects made her feel freer to choose what she liked, particularly regarding finding information about a given topic. Overall, they all agreed that carrying out a project as an assessment procedure motivated them because they were more interesting and entertaining, as John remarked “...*things like the fortune teller project motivated me because it was entertaining and innovative*” [free translation]. Thus, the results reported herein seem to corroborate previous studies (Petre, 2017) in that the use of alternative assessment procedures facilitate and enhance language motivation levels in students as they need authentic, dynamic, innovative, and creative tasks, which will eventually lead to higher motivation levels.

Anxiety has been acknowledged as one of the major barriers to learning a foreign language. During the focus group interview, Anna reported that that she felt more anxious when having to take a test because all the tension was concentrated on a particular day, which brings about higher levels of anxiety. Anxiety is often disruptive and leads to performance decrements in evaluative situations. By the same token, the use of alternative assessment procedures can have a positive effect on reducing anxiety level of students as projects - as was the case of this study - provide students with a more neutral and less stressful way of demonstrating what they have learned by doing something practical (Al-Ruqeishi & Al-Humaidi, 2016).

The perceived gains of utilising an alternative assessment procedure also include the potential for a more varied strategy use. Indeed, Jenny argued that the use of the project “...*allows me to check the con-*

*tents in my notebook to generate more ideas...*” [free translation]. This notion is shared by other participants and is further expanded by Paul, who claimed that he felt at ease when working on projects because he had the chance to check his work during the process and make the ensuing corrections. Similarly, John reported that when working on projects “*one has the chance to write and send it out for review; then I check and correct it again. Then I’m sure it is correct because you checked it earlier*” [free translation]. In other words, as put by Anna, students “*have greater freedom*” when carrying out an alternative assessment procedure.

Somewhat aware of individual differences in learning, John also underscored that by using projects for assessment purposes, everybody could show what they knew and that “*... there are some people who express themselves in different ways, some classmates write, others speak, or even draw, so this makes learning easier ... so there are different types of abilities the students can use*” [free translation]. Finally, on the same note, Peter stated that “*it is easier for me because there are different forms of learning apart from studying* [for a test]”.

All in all, it can be safely argued that the qualitative findings reported above confirm what Libman (2010) stated in that alternative assessment procedures go beyond traditional psychometrically driven testing as these kinds of assessment procedures require students to produce and demonstrate knowledge rather than recalling information (p. 63).

## Conclusion

This study has examined the washback effect of using alternative assessment procedures and traditional tests on a seventh grade class in a subsidized school in Chile, and attempted to unpack students’ perceptions of the impact of administering different assessment and testing procedures on their motivation, anxiety and strategy use.

Although there was no statistical difference in the perceived washback effect of the two assessment procedures, it can be safely claimed

that, the use of the alternative assessment procedure - a project as was the case in this study - produced a comparatively more positive washback effect, a finding that corroborates previous studies. Hung (2012), for instance, found that the use of alternative assessment procedures did indeed generate positive washback effects on learning by 'building a community of practice, facilitating peer learning, enhancing learning of content knowledge, and cultivating critical thinking' (p.33), which attests to the alleged superior positive washback effect of alternative assessment procedures.

It must be noted that the use of traditional assessment procedure in the sample did not yield a downright negative washback effect, which - at least on the surface - seems to go counter to what the literature suggests (Haggerty & Fox, 2015). Indeed, the results suggested that testing need not be perceived as a negative assessment procedure capable of producing purely negative feelings in the learners; nonetheless, it can be argued that even in contexts where learners value positively the use of traditional tests, probably due to other extraneous variables, the use of alternative assessment procedures still enjoys greater learners' acceptance and is perceived as non-threatening, thus boosting confidence and learner strategy use as it constitutes an opportunity for students 'to take more responsibility for their own learning by getting them involved in activities that stimulate student's abilities to create and apply a wide range of knowledge rather than engaging in acts of memorization and basic skills development' (p. 63).

One of the limitations that needs to be acknowledged deals with the point in time at which the study was conducted. As the school year was drawing to a close, the students' attention was not particularly focused. Also, during the focus group interview, two of the participants came across as rather bashful and provided particularly short responses, while other participants attempted to dominate the discussion, an issue that most likely relates to the maturity level of the participants. Thus, it can be suggested that other complementary data-gathering instruments could be used to gain a better understanding of the perception of washback effect.

## References

- Alcaraz, N. (2015). Aproximación histórica a la evaluación educativa: De la generación de la medición a la generación ecléctica. *Revista Iberoamericana de Evaluación Educativa*, 8(1), 11-25.
- Armstrong, R. A. (2014). When to use the Bonferroni correction. *Ophthalmic & Physiological Optics*, 34(5), 502-508.
- Akpınar, K. D. & Çakıldere, B. (2013). Washback effects of high-stakes language tests of Turkey (KPDS and ÜDS) on productive and receptive skills of academic personnel. *Journal of Language and Linguistic Studies*, 9(2), 81-94.
- Alderson, J., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14, 115-129.
- Alderson, J. C., Brunfaut, T., & Harding, L. (2017). Bridging assessment and learning: a view from second and foreign language assessment. *Assessment in Education: Principles, Policy & Practice*, 24(3), 379-387.
- Al-Ruqeishi, M., & Al-Humaidi, S. (2016). Alternative Assessment as Perceived by EFL Teachers. *IUP Journal of English Studies*, 11(3), 88-101.
- Aydin, S. (2010). EFL writers' perceptions of portfolio keeping. *Assessing Writing*, 15(3), 194-203.
- Brown, H. D. (2004). *Language assessment: Principles and classroom practices*. San Francisco, CA: Pearson ESL.
- Cheng, L., & Curtis, A. (2012). Test impact and washback: Implications for teaching and learning. In C. Coombe, P. Davidson, B. O'Sullivan, & S. Stoyhoff (eds.), *Cambridge guide to Second Language Assessment*, 89-95. New York: Cambridge University Press.
- Cheng, L., & DeLuca, C. (2011). Voices from Test-Takers: Further Evidence for Language Assessment Validation and Use. *Educational Assessment*, 16(2), 104-122.
- Cheng, L., Sun, Y., & Ma, J. (2015). Review of washback research literature within Kane's argument-based validation framework. *Language Teaching*, 48(4), 436-470.
- Collins, S., Reiss, M., & Stobart, G. (2010). What happens when high-stakes testing stops? Teachers' perceptions of the impact of compulsory national testing in science of 11-year-olds in England and its abolition in Wales. *Assessment in Education: Principles, Policy & Practice*, 17(3), 273-286.
- Contreras, A., Veliz-Campos, M., L., Veliz (2019). Portfolios as a strategy to lower English language test anxiety: The case of Chile. *International Journal of Instruction*, 12(1), 181-198.

- Creswell, J. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches*. London: SAGE Publications, Incorporated.
- Fox, J. (2017). Using Portfolios for Assessment/Alternative Assessment. In Shohamy E., Or I., May S. (eds), *Language Testing and Assessment. Encyclopedia of Language and Education*, 135-148. Auckland: Springer.
- Fulcher, G. (2010). *Practical language testing*. London, UK: Hodder Education.
- Ghoorchaei, B., Tavakoli, M., & Ansari, D. N. (2010). The impact of portfolio assessment on Iranian EFL students' essay writing: A process-oriented approach. *GEMA Online® Journal of Language Studies*, 10(3), 35-51.
- Haggerty, J. F., & Fox, J. (2015). Raising the bar: Language testing experience and second language motivation among South Korean young adolescents. *Language Testing in Asia*, 5, 1-16.
- Hussain, N., Nawas, B., Nasir, S., Kiani, S., & Hussain, N. (2013). Positive teacher, relationship and teacher's experience: A teacher's perspective. *Global Journal of Management and Business Research Interdisciplinary*, 13(3.1), 1-4. Disponible en: [https://globaljournals.org/GJM\\_BR\\_Volume13/1-Positive-Teacher-Student-Relationship.pdf](https://globaljournals.org/GJM_BR_Volume13/1-Positive-Teacher-Student-Relationship.pdf)
- Hughes, A. (2003). *Testing for language teachers*. Cambridge, UK: Cambridge University Press.
- Hung, S. (2012). A washback study on e-portfolio assessment in an English as foreign language teacher preparation program. *Computer Assisted Language Learning*, 25(1), 21-36.
- Inbar-Lourie, O., & Donitsa-Schmidt, S. (2009). Exploring classroom assessment practices: The case of teachers of English as a foreign language. *Assessment in Education: Principles, Policy & Practice*, 16(2), 185-204.
- Lam, R. (2014). Promoting self-regulated learning through portfolio assessment: Testimony and recommendations. *Assessment & Evaluation in Higher Education*, 39(6), 699-714.
- Libman, Z. (2010). Integrating Real-Life Data Analysis in Teaching Descriptive Statistics: A Constructivist Approach. *Journal of Statistics Education*, 18(1). DOI: 10.1080/10691898.2010.11889477
- Ma, J., & Chen, L. (2016). Chinese students' perceptions of the value of test preparation courses for the TOEFL iBT: Merit, worth and significance. *TESL Canada Journal*, 33(1), 58-79.
- McNamara, T. (2000). *Language testing*. Oxford: Oxford University Press.
- Meckes, L., & Carrasco, R. (2010). Two decades of SIMCE: an overview of the National Assessment System in Chile. *Assessment in Education: Principles, Policy & Practice*, 17(2), 233-248.
- Meyer, M. A., & Booker, J. M. (2001). *Eliciting and analyzing expert judg-*

- ment: a practical guide*. Philadelphia: Society for Industrial: Applied Mathematics and American Statistical Association.
- Petre, A. L. (2017). The impact of Alternative Assessment Strategies on students. *Scientific Research & Education in the Air Force-AFASES*, 2, 157-160.
- Polesel, J., Rice, S., & Dulfer, N. (2014). The impact of high-stakes testing on curriculum and pedagogy: A teacher perspective from Australia. *Journal of Education Policy*, 29(5), 640-657.
- Purpura, J. E. (2016). Second and foreign language assessment. *The Modern Language Journal*, 100(1), 190-208.
- Safa, M. A., & Goodarzi, S. (2014). The Washback Effects of Task-based Assessment on the Iranian EFL Learners' Grammar Development. *Procedia-Social and Behavioral Sciences*, 98(14), 90-99.
- Shawn, P., & McMillion, A. (2008). Proficiency effects and comprehension in advanced second language reading. *Nordic Journal of English Studies*, 7(3), 123-143.
- Shohamy, E. G. (2001). *The power of tests: A critical perspective on the uses of language tests*. New York: Pearson Education.
- Shohamy, E., Or, I., & May, S. (2017). *Language Testing and Assessment. Encyclopedia of Language and Education (3rd ed.)*. Auckland: Springer.
- Tayeb, Y. A., Aziz, M. S. A., Ismail, K., & Khan, A. B. M. A. (2014). The Washback Effect of the General Secondary English Examination (GSEE) on Teaching and Learning. *GEMA Online Journal of Language Studies*, 14(3), 83-103.
- Tsagari D., & Cheng L. (2017). Washback, Impact, and Consequences Revisited. In Shohamy E., Or I., May S. (eds), *Language Testing and Assessment. Encyclopedia of Language and Education*, 359-372. Auckland: Springer.
- Williams, P. (2014). Squaring the circle: A new alternative to alternative assessment. *Teaching in Higher Education*, 19(5), 565-577.
- Xie, Q., & Andrews, S. (2013). Do test design and uses influence test preparation? Testing a model of washback with Structural Equation Modeling. *Language Testing*, 30(1), 49-70.
- Zhan, Y. & Andrews, S. (2014). Washback effect from a high-stake examination on out-of-class English learning: insight from possible-self theories, *Assessment in Education: Principles, Policy & Practice*, 21(1), 71-89.
- Zhao, L. (2015). The influence of learners' motivation and attitude on second language teaching. *Theory and Practice in Language Studies*, 5(11), 2333-2339.

Recibido: 15.05.2019. Aceptado: 15.10.2019