



Universidad Austral de Chile  
Facultad de Filosofía y Humanidades  
Instituto de Lingüística y Literatura

Lead advisor:  
MS. Andrea Lizasoain

**English Teachers' Assessment Practices in Valdivia in the light of the Principle of  
Reliability**

Seminario de Tesis para optar al Título de Profesora en Comunicación en Lengua Inglesa y  
al Grado de Licenciada en Educación

Andrea Mara Ramos Órdenes

Valdivia, Chile

2013

English Teachers' Assessment Practices in Valdivia in the light of the Principle of Reliability.

**ANDREA MARA RAMOS ÓRDENES**

### **Acknowledgments**

To my friends, who have always been there at the most difficult and beautiful moments of my life.

To Carla, who has always given her unconditional support throughout this entire process.

To Mrs. Andrea Lizasoain, who has been a fundamental part of this enterprise.

To the English teachers who made this research possible.

*“Work for a cause, not for applause. Live life to express and not to impress. Do not strive to make your presence noticed, just make your absence felt”* (Unknown Author).

## TABLE OF CONTENTS

Acknowledgments.....	i
Abstract.....	1
Introduction .....	2
1.CHAPTER I: Theoretical Background.....	4
1.1. English Teaching in Chile.....	4
1.1.1. English as a Foreign Language in Chile .....	4
1.1.2. English Opens Doors Program .....	5
1.1.3. CAMBRIDGE and SIMCE results.....	6
1.2. Antecedents of the study .....	7
1.2.1. Research question .....	7
1.2.2. Hypothesis.....	8
1.2.3. Objectives.....	8
1.2.3.1. General objective .....	8
1.2.3.2. Specific objectives.....	8
1.3.A Definition of Evaluation, Assessment and Testing.....	8
1.3.1. Evaluation.....	8
1.3.2. Assessment.....	9
1.3.2.1. Formal and informal Assessment.....	10
1.3.3. Benefits of Assessment in the English Classroom.....	11
1.3.3.1. Assessment Benefits for Teachers .....	11
1.3.3.2. Assessment Benefits for Students .....	13
1.3.4. Testing.....	16

## English Teachers' Assessment Practices in Valdivia in the light of the Principle of Reliability.

1.4. The Principles of Assessment.....	17
1.4.1. Validity.....	18
1.4.2. Authenticity .....	19
1.4.3. Practicality.....	20
1.4.4. Backwash.....	21
1.4.5. Reliability .....	22
1.4.5.1. Student –related Reliability.....	22
1.4.5.2. Rater Reliability .....	23
1.4.5.2.1. Inter-rater Reliability.....	24
1.4.5.2.2. Intra-rater Reliability.....	24
1.4.5.3. Test Administration Reliability.....	27
1.4.5.4. Test Reliability .....	30
2. CHAPTER II: Methodology .....	32
2.1. Justification .....	32
2.1.1. Assessment in Latin America.....	32
2.1.2. English Assessment in Chile .....	32
2.2. Educational Survey Research.....	33
2.3. Data Collection Method .....	34
2.4. Data Collection Analysis.....	34
2.5. Subjects.....	35
3. CHAPTER III: Results .....	38
4. CHAPTER IV: Discussion .....	45
4.1. Intra-rater Reliability .....	45

English Teachers' Assessment Practices in Valdivia in the light of the Principle of Reliability.

4.2. Test Administration Reliability .....	47
Conclusion .....	53
References.....	55
Appendixes .....	66

## **English Teachers' Assessment Practices in Valdivia in the light of the Principle of Reliability**

### **Abstract**

This current educational research aims to portray the assessment practices of English Foreign Language (EFL) teachers in the Commune of Valdivia. This study was focused on two elements of the principle of reliability: Intra-rater reliability and test-administration reliability, proposed by Brown (2004). In order to do so, 26 English teachers from 19 educational settings participated by answering a survey during the second semester of 2012. Results showed that the intra-rater reliability principle was only partially met due to one element that was not completely fulfilled. However, test administration reliability was totally met.

**Key words:** *Assessment practices, EFL, principles of assessment, intra-rater reliability, test administration reliability*

## **Introduction**

English teaching has not been effective in Chile in spite of the Governments' efforts. Results from diagnostic tests have shown that Chilean students have not reached international standards of proficiency in English (Red Maestros de Maestros, 2004). In relation to this, findings demonstrated that students from private institutions obtained better outcomes than those in public and subsidized high schools (Arango, 2008). In Matear's (2008) words, the main problem has been the fact that English lessons are carried out in Spanish and are grammar-based (as cited in Gutierrez & Peña, 2011). These facts have affected the way in which English has been taught, producing low English results and performance. It has also been found that the primordial role of English in Chile is primarily academic and economic rather than communication (Palabra Maestra, 2011). However, nowadays, the Chilean Ministry of Education has emphasized the importance of language communication through initiatives for both EFL educators and pupils in public and subsidized sectors (Palabra Maestra, 2011; EducarChile, 2011).

The native speaker ex-counselor of the Ministry of Education in terms of English matters, Mr. Andrew Sheehan, has said that the poor results of the Chilean population in English proficiency might be related to the lack of testing expertise in the country. In fact, it is common to find data about English Second Language (ESL) testing rather than English as a foreign language in many countries. However, there is little information related to testing in Chile. The lack of assessing preparation and the unawareness of the principles of assessment has led to teachers working without a scientific basis (McNamara, 2004; Figueras, 2005). This point is relevant as, although assessment is an inherent activity in the teaching praxis, it has been underestimated in the teaching field. Regarding this, H.

English Teachers' Assessment Practices in Valdivia in the light of the Principle of Reliability.

Douglas Brown (2004) has proposed five principles of language assessment: Validity, authenticity, practicality, washback and reliability; which would contribute largely to the overall teaching practice if teachers in general knew them and put them into practice.

Nevertheless, not even the Ministry of Education put these principles into practice when they apply proficiency tests (2004, 2008, 2010, and 2012), such as the Cambridge ESOL Examinations and the TOEC tests, to diagnose the general level of English language competence of the Chilean people. This reality clearly reflects the lack of relevance assessment is given in the country.

In order to both contribute to the research field regarding assessment and help improve assessment practices by teachers in the country, an educational survey research study was designed. Twenty-six English Foreign Language (EFL) teachers participated in order to diagnose the assessment practices in the Commune of Valdivia, in the light of two elements of the principle of reliability described by Brown (2004): Intra-rater and test administration reliability have been analyzed throughout this paper.

The present paper has been organized as follows: Chapter I depicts the state of the art regarding assessment practices in Chile and defines baseline concepts, such as evaluation, testing and assessment; benefits of assessment in the English classroom; and the principles of assessment. Chapter II describes the methodology designed in order to find out whether teachers in Valdivia meet the principle of reliability or not. Chapter III shows the results obtained in the light of the principle. And, finally, Chapter IV presents the discussion of the results.



## **1. CHAPTER I: Theoretical Background**

### **1.1. English Teaching in Chile**

#### **1.1.1. English as a Foreign Language in Chile**

Before giving a picture of the Chilean reality regarding English teaching, the terms English as a Foreign Language and English as a Second Language must be clarified since there is some misunderstanding between these concepts. On the one hand, it can be said that English as a Second Language (ESL) is the language which is not the mother tongue or first language (L1), but it is the language with which individuals are constantly surrounded or exposed to, because it is the official language of a country. On the contrary, English as a Foreign Language (EFL) is not the official language of the community neither of the school; it is a language which is taught restrictively in the classroom under teaching environments (Gunderson, 2008).

The importance of the English language yields in the fact that, it is the most popular foreign language taught in Chile and around the world. English has become a global language and, as David Crystal (2003) states, “a language achieves a genuinely global status when it develops a special role that is recognized in every country” (p. 3). Its use has been throughout mass media, politics, education and technology, among others. As a global language, English has been promoted in non-English speaking countries such as the Chilean case.

According to the English Opens Doors Program ex Director, Rodrigo Fábrega (2011), English in Chile has been taught for at least 250 years<sup>1</sup>. In this sense, it is the most frequent foreign language taught in Chilean schools being part of the national educational curriculum. It has continually taken part of Chilean life providing students with new tools

---

<sup>1</sup> All translations in this paper were carried out by the author

English Teachers' Assessment Practices in Valdivia in the light of the Principle of Reliability.

to enhance their job and educational opportunities (Palabra Maestra, 2011). However, international results of language tests have demonstrated that in spite of the efforts, Chile has obtained negative outcomes in comparison to other countries in South America. The Educational First English Proficiency Index [EF EPI] (n.d), an organism created to measure, compare, and standardize the level of English in non-English speaking countries, gathered scores for three years using online tests applied to more than 2 million adults around the world. Results yielded that South American countries have the lowest level of English worldwide, being the main cause of this the low quality of public education (*ibib*). It is essential to mention that this study did not consider Uruguay as part of South American countries to be studied. In the case of Chile, the score proves that Chileans adults (very low level; 44.63 EF EPI) scored a “very lower level of English” (p.16, n.d.) in comparison to adults in Argentina who obtained intermediate level with 53.49 EF EPI results; following Brazil with low level and marking 47.27 EF EPI, Peru with very low level and 44.71 EF EPI results. In this light, it can be said that Argentineans, Brazilians and Peruvians scored higher than the Chilean counterpart. However, Chile obtained better results than Ecuador that has very low of English and 44.54 EF EPI, Venezuela with a very low level with 44.43 EF EPI, and Colombia with a very low level of English with 42.77 EF EPI (*ibib*).

### **1.1.2. English Opens Doors Program**

In 2003, the Chilean English Opens Doors Program was created in order to help all Chilean students to reach an instrumental level of English, which would broaden their educational and employment opportunities. The main objective of this program was to increase the country's economic development (Palabra Maestra, 2011). The plan promoted the use of English as a fundamental part of socio-cultural contexts as well as employment

English Teachers' Assessment Practices in Valdivia in the light of the Principle of Reliability.

opportunities (EducarChile, 2005). As Crystal (2003) acknowledges, when a language has an official status is part of business, formal governmental speeches, access to privilege academic information and access to worldwide job opportunities. That is to say, there is a strong relationship among language dominance and economic, technological, and cultural power “without a strong power-base, of whatever kind, no language can make progress as an international medium of communication” (Crystal, 2003, p. 7).

The Chilean English Opens Doors Program stated that by 2013, Chilean students would have reached the KET (Key English Test) level when finishing primary education, and PET (Preliminary English Test) level when ending high education. In order to achieve these objectives, the program focused on improving four main areas: (a) the use of international standards, (b) English teachers' professional development, (c) English learning support for schools, and (d) employment opportunities (Palabra Maestra, 2011).

### **1.1.3. CAMBRIDGE and SIMCE results**

As a starting point, the MINEDUC, through the English Opens Doors program, asked Cambridge ESOL Examination to develop an English proficiency test that was specially addressed to the Chilean reality. The test was administered in 2004 to 11,000 students in 8<sup>th</sup> and 12<sup>th</sup> grades from 299 different schools. Results evidenced that EFL students could not reach the Threshold/ALTE 2 level. Moreover, a low percentage of Chilean students could only achieve the Waystage/ALTE1 level and the majority of them were only able to reach the Breakthrough, Lower Breakthrough and Pre-Breakthrough levels (Red Maestros de Maestros, 2004). Even though, those levels were created taking into consideration the Chilean reality the results were negative. Reason why, the ex Director of the English Opens Doors Program implemented new strategies to improve the English level of Chilean students. Among these strategies the following can be found: (a) the use of the English

English Teachers' Assessment Practices in Valdivia in the light of the Principle of Reliability.

learning theories, (b) partnerships with English speaking countries, and (c) the continuance of the steps to achieve long-term goals in spite of governmental changes (Palabra Maestra, 2011).

In 2010, the English SIMCE test was applied for the first time in Chile and Latin America. The test was developed this time by the Educational Testing Service (ETS). This instrument was the Test of English for International Communication (TOEIC Bridge), which focused on the receptive skills (listening and reading) of the language, under the international standards of the Association of Language Testers in Europe (ALTE). It was applied to 240,000 EFL eleventh graders from subsidized, semi-subsidized and private high schools (EducarChile, 2011).

According to EducarChile (2011), the results obtained through the English SIMCE showed that Chilean students had not reached the minimal international standards. In other words, to achieve a basic level of English, students must score at least 134 points. Unfortunately, EFL Chilean learners averaged only 99 points, which means that only 11% of students obtained a certification and 89% could not reach a basic international level of English. This means that, EFL Chilean students could barely communicate using English, but the one who reached the basic level could understand short phrases and common expressions (listening comprehension) and could comprehend short and simple reading texts (reading comprehension) (*ibid.*).

## **1.2. Antecedents of the study**

### **1.2.1. Research question**

The following research question addresses the importance of assessment as part of the teaching and learning practices: Do English teachers in Valdivia meet the assessment principle of reliability at the time of assessing their students (Brown, 2004)?

### **1.2.2. Hypothesis**

English teachers in Valdivia meet the principle of reliability when assessing their students' English skills.

### **1.2.3. Objectives**

#### **1.2.3.1. General objective**

This study seeks to determine if EFL teachers meet the principle of reliability when assessing students' performance.

#### **1.2.3.2. Specific objectives**

This research study aims at finding out whether teachers in Valdivia meet one of elements of the principle of intra-rater reliability and to see if test-administration reliability is also met. That way, it will be possible to diagnose general assessment practices in the city and the country.

### **1.3. A Definition of Evaluation, Assessment and Testing**

Evaluation, assessment and testing are commonly used terms in educational contexts. However, these three concepts are usually misunderstood and considered synonyms. Therefore, the following paragraphs provide the reader with a further explanation of these three terms in order to elucidate them.

#### **1.3.1. Evaluation**

Evaluation in educational settings involves both assessment and testing practices (Brown, 2004). In this regard, it can be stated that evaluation embraces both assessment and testing which are fundamental components of evaluation itself. Christine Coombe, Keith Folse, and Nancy Hubley (2007) report, "evaluation is all-inclusive and is the widest basis for collecting information in education" (p. xv). That is, evaluation is a means of systematic investigation in which observation and analysis of the collected information are the grounds

English Teachers' Assessment Practices in Valdivia in the light of the Principle of Reliability.

to obtain results. These results can enhance teaching practices and can be used to reconsider the usefulness of learning activities. Consequently, evaluation is “not only assembling information but interpreting that information – making value judgements” (Nunan, 1992, p. 185). Following this, evaluation gathers the information needed in order to improve a course in all its aspects. In this regard, Michael Harris and Paul McCann (1994) mention, “evaluation involves looking at all factors that influence the learning process, .i.e., syllabus, objectives, course design, and materials” (p. 2). Hence, evaluation has an impact on students' way of learning as well as on teachers' methodologies.

### **1.3.2. Assessment**

As Coombe *et al.* (2007) state, “assessment is an integral part of the teaching/learning cycle. It involves planning, development, administration, analysis, feedback, and reflection” (p. 13). Therefore, assessment is an essential component of the curriculum and teaching practices. As it is exemplified by Graham Butt (2010) “assessment should be placed at the very centre of the act of teaching and learning – our ideas about how students learn, how we plan lessons, how we choose our assessment methods and our concepts of intelligence should all connect” (p. 13). In this sense, Butt (2010) suggests, assessment must include educational goals and clear objectives. The importance of having clear objectives at the time of teaching makes assessment more productive and efficient. Therefore, this helps educators to see EFL learners' improvement and analyze this information to apply it in the classroom.

As it can be illustrated, assessment portrays the way in which teachers promote learning and it is basically inherent to teachers; “a teacher never ceases to assess students whether those assessments are incidental or intended” (Brown, 2004, p. 4). In this manner, assessment is a holistic process in which teachers as well as pupils work together, which

English Teachers' Assessment Practices in Valdivia in the light of the Principle of Reliability.

means that in order to meet objectives, improve teaching/learning practices and students' results. Educators must think about lesson planning, assessment methods, administration of tests and differentiate the way with students learn new concepts to enhance teaching practices.

#### **1.3.2.1. Formal and informal Assessment**

Historically, assessment has been divided into two categories: formal and informal (Brown, 2004). On the one hand, formal assessment is a “systematic planned sampling technique constructed to give teacher and student an appraisal of student achievement” (Brown, 2004, p. 6). It is scheduled and periodic, and it helps educators and students to be aware of their weaknesses and strengths during the entire course. On the other hand, informal assessment, it is not scheduled and can occur at any instance in the classroom. As Harris and McCann (1994) observe, it is a method of collecting information about students' performance in ordinary classroom conditions. In the same line, H. Douglas Brown (2004) states, informal assessment elicits students' performance without recording results. In this regard, it promotes students' adaptability in order to be more familiar with assessment procedure as well as make learners use and reinforce the contents learnt.

To conclude, formal assessment is a frequent means of obtaining feedback about students and teachers' performance in order to obtain concrete quantifiable results of the progress made so far in the learning process (Brown, 2004; Coombe *et al.*, 2007; Richard-Amato, 2005); while informal assessment is constantly happening and does not focus specifically on results, but on progress. It seeks to foster student's abilities in the language and it helps students to be familiar with different assessment procedures (Harris & McCann, 1994), which is only part of the benefits assessment has to offer.

### **1.3.3. Benefits of Assessment in the English Classroom**

The benefits of assessment techniques have influenced both teachers and learners (Brown, 2004; Coombe *et al.*, 2007; Organization for Economic Co-Operation and Development [OECD], 2005; Richard-Amato, 2005). Assessment methods encourage long term goals, thus improving the teaching/learning practices. That is to say, they promote the link between classroom assessments and students' learning (Richard-Amato, 2005). In particular, these techniques provide teachers with essential data with which they can identify students' strengths and weaknesses, can monitor educatees' learning and progress, and can plan and conduct class instruction (Educational Testing Service, 2003).

#### **1.3.3.1. Assessment Benefits for Teachers**

Assessment provides teachers with opportunities to evaluate their own performance (Brown, 2004; Coombe *et al.*, 2007; OECD, 2005; Richard-Amato, 2005). The results of the assessments can reveal weaknesses or strengths in terms of teaching instruction (Gordon, Kane & Staiger, 2006). In consequence, educators can use the results of assessments to change or improve the curriculum and the delivery of class content, taking more or less time to develop them (Nicol, 2007). By doing this, not only classroom teachers help pupils, but they also improve their effectiveness as educators (Hanushek, Kain, & Rivkin, 2005).

- **Culture of the Classroom**

In order to enhance effective teachers' praxis, it is necessary to foster a positive learning environment. That is, culture of the classroom, which is defined as a positive learning environment (Butt, 2010; Richard- Amato, 2005; Ruhul & Wannarka, 2008). This is promoted when educators help learners take risks and make mistakes to boost students'



English Teachers' Assessment Practices in Valdivia in the light of the Principle of Reliability.

self-confidence and have a positive view of the teaching environment (Brown, 2007; OECD, 2005). In addition, a positive culture of a classroom is encouraged when educators integrate students with different backgrounds in their activities, because this enhances respect among the pupils (Barker, Farrelly & Frederiks, n.d). In this way, teachers “create a classroom that will inspire student learning by making students feel energized and empowered” (Point Loma Nazarene University, 2012, para. 1).

Cultural interactions foster a non-threatening environment in the classroom (Aslett, 2006; Popham, 2012). As Michael Byram and Anwei Feng (2005) state, “recent research studies have shown the beneficial effects of incorporating cultural awareness in language classrooms” (as cited in Brown, 2007, p. 200). The integration of cultural aspects is relevant since educators need to address activities in order to encourage students' positive view of the English language and so, to be successful educators (Saville-Troike, 1978).

- **The Efficient Use of Feedback**

Feedback allows students to correct conceptual errors and encourages educators to modify instructional activities in the light of their effectiveness (Butt, 2010; Spiller, 2009). Jean Mandernach (2003) supports this by adding that “assessments are designed to guide learning and are not utilized as an outcome measure” (para.1). In order to do so, feedback has to be a timely and specific tool that students use to know what it is expected from them (Coombe *et al.*, 2007; Gedy, n.d). As David Nicol (2007) points out, educators must “deliver high quality feedback information that helps learners self-correct” (p. 3). Consequently, they can track students' progress, change activities and reinforce contents (*ibid.*). “With assessment, the instructor can assign a grade, but also [the instructor] explores how to improve student learning” (The University of Texas, 2011, para. 2). In this

English Teachers' Assessment Practices in Valdivia in the light of the Principle of Reliability.

line, the use of effective feedback facilitates the students' learning process as well as the educators' job.

- **Assessing Learners' Knowledge**

In the teaching field, it is relevant to consider the role of assessing pupils' understanding. In this light, Tina Blythe and Associates (1998) mention that, at the moment of assessing learners' knowledge, educators must consider: Goals, methods to check understanding, criteria to assess students, and feedback. Hence, they can reinforce contents, change activities, ensure grading criteria and to improve teaching instruction (Texas University, 2011).

Assessing learners' understanding improves educators' praxis as well as students' test results (Woytek, n.d). However, educators must assess students continually, in order to obtain benefits. In this regard, Patricia Richard-Amato (2005) states that "assessment at its best can become an integral part of what happens in the classroom" (p. 151). Assessment, then, is not only grading students' tests, in fact it is an overall process (Brown, 2004; Butt, 2010; Coombe *et al.*, 2007; Richard-Amato, 2005).

### **1.3.3.2. Assessment Benefits for Students**

The benefits of assessment techniques promote learners' accomplishment of knowledge rather than a testing-fragmented part of knowledge (Brown, 2004). This overall view of assessment is the basis for teachers to design a syllabus that includes many tools to gather information about students (Baldwin, McFadyen & Stringer, 2010). By doing this, educators enhance a student-centered approach to teach that often leads to student success (OECD, 2005).

- **Educational Achievement**

Achievement gains associated to assessment have been found when learners are involved in their own learning process (Brown, 2007; Haugen, 1999; Richard-Amato, 2005). This is enhanced when educators address students' individual needs and assess students' learning (Butt, 2010; Richard-Amato, 2005). However, in order to achieve gains not only do teachers play an important role, but also students. In this light, Brown (2007) mentions that learning a foreign language requires motivation to achieve goals. Richard-Amato (2005) adds that motivation involves temporary expectancies, curiosity, ego and personal satisfaction among others. Additionally Coombe *et al.* (2007) emphasize that students need to have a study plan and clear goals to progress. Otherwise, learners will not achieve better results. Hence, students need to be self-motivated to accomplish their goals. Additionally, educators need to assess students' progress. In Butt's (2010) words, "student achievement needs to be accurately measured at the start of the process and then re-assessed, once interventions have been made" (p. 85). Learners can be an important part of their learning only if they are in charge of it. However, educators need to track students' progress to motivate them (Butt, 2010). This is the only way students can know their weaknesses and strengths so they are aware of what they need to improve: Their goals.

- **Educational Equity**

According to Jittie Brandsma (n.d), "equity refers to "justice" and forms essentially a basis or justification for government interventions in education" (p.15). In this regard, assessment has proved to improve equity of students' outcomes (OECD, 2005). Aspects such as attendance, retention of learning and an improvement in the quality of learners' tasks demonstrate these findings (Coombe *et al.*, 2007; OECD, 2005). It has been found

English Teachers' Assessment Practices in Valdivia in the light of the Principle of Reliability.

that students who usually performed poorly increased their learning tremendously after the intervention of meaningful assessment strategies.

Reasons for inequity assessment are exclusion (Brandsma, n.d; Gamoran, 2002), segregation (Gamoran, 2002), racism (Weissglass, 2003) and socio economic status (Jackson, Smith & Sparks, 2003). Additionally, Butt (2010) identifies five reasons for inequity: Lack of motivation through the contents taught, type of teaching experienced, bias towards students, performance expectations, and poor or lack of feedback. These factors have a negative impact on assessment and promote inequity in the tests results (Butt, 2010; Weissglass, 2003). Therefore, in order to foster equity in the assessment process, educators must objectively assess students in spite of their differences.

- **Learning Autonomy**

According to Anna Chamot (2005), strategies are “procedures that facilitate a learning task... Strategies are most often conscious and goal driven” (as cited in Brown, 2007, p. 132). In this light, students need to assess their own learning process. As Brown (2007) points out, “closely linked to the concept of autonomy is the demand on learners to become aware of their own process of learning” (p. 131). Students using learning strategies “formulate their own learning goals, track their progress on achieving these goals, identify their knowledge deficits, detect contradictions,... [and] make inferences (Graesser, 2011, para. 3). Learners choose strategies based on “their learning styles and preferences, personalities, and cultural backgrounds” (Richard-Amato, 2005, p. 84). In this respect, Lee Haugen (1999) states that development of efficient study techniques increase the ability to think critically about the course content and foster an attitude that values long-term retention.

Learning to learn skills benefits students during their entire life (Coombe *et al.*, 2007). In order to do so, pupils need to become self-directed learners. That is, “students must learn to monitor and adjust their approaches to learning” (Carnegie Mellon University, n.d, para. 7). To finish, it can be stated that learning skills and autonomy are interwoven in the process of learning (Brown, 2007; Chamot, 2005; Richard-Amato, 2005).

#### **1.3.4. Testing**

According to Brown (2004), testing is defined as “a method of measuring a person’s ability, knowledge, or performance in a given domain” (p. 3). It is basically the use of tests as instruments to gather information on students’ knowledge of a specific topic and on a field of teaching. That is, tests are the instruments to measure learners’ abilities and current level of knowledge in a more concrete manner. In this way, testing is a branch of assessment categorized as a formal means of eliciting results. In this regard, Brown (2004) mentions, “all tests are formal assessment, but not all formal assessment is testing” (p. 6). For example, the use of a journal or a portfolio is not considered as a part of the testing category, but they can be formal assessment. Hence, testing is the action of tests application, which yields quantitative results based on the students’ current level of knowledge.

Evaluation, assessment and testing are a fundamental part of educational settings (Brown, 2004; Butt, 2010, Coombe *et al.*, 2007). Educators and pupils are continually facing diverse aspects of them in the language courses throughout the entire semester. Evaluation refers to the use of data in order to analyze teaching and learning progress. It is the wide-ranging area which includes both assessment and testing (Coombe *et al.*, 2007; Harris & McCann, 1994; Nunan, 1992). In other words, evaluation is an all inclusive process which includes aspects of assessment such as planning, development,

English Teachers' Assessment Practices in Valdivia in the light of the Principle of Reliability.

administration, analysis, feedback and reflection. Additionally, it incorporates aspects of testing such as the use of tests as formal tools to attain results measuring learners' knowledge in a specific matter. Finally, assessment and testing are important components of evaluation without them evaluation could not exist by itself.

#### **1.4. The Principles of Assessment**

Teachers and testers should design test items following the principles of language assessment stated by Brown (2004): Validity, authenticity, reliability, practicality, and washback. The importance of following the principles of language assessment lies in the fact that, they are tools to construct well-designed classroom tests with which educators are able to obtain objective information about students' learning (*ibid*).

The principles of language assessment have been studied by numerous researchers and applied by educators and testers. This fact demonstrates that there is some agreement among them. The following section includes the considerations proposed by Coombe *et al.* (2007), and Penny Ur (1999).

Penny Ur (1999) includes the principles of assessment as part of three stages of test design: test preparation, performance and feedback. The first stage includes: a) validity, which is related to test items and its relation to test content; b) clarity, which refers to the use of clear instructions for each item; c) do- ability, which means to create questions neither difficult nor tricky; d) marking, which regards the use of clear rubrics, interesting design tests with relevant topics; and e) heterogeneity, which deals with the inclusion of easy and difficult items. The second stage, performance, it is when a teacher applies a test to other students before giving it to their course; and the third stage is about feedback the teacher must ask students their opinions about the test and how they felt about the test. In contrast, Coombe *et al.* (2007) propose the following as corner stone of testing: a)

English Teachers' Assessment Practices in Valdivia in the light of the Principle of Reliability.

usefulness, which refers to the specific purpose, the testees and the language in a test; b) transparency, which deals with the information that teachers or testers should give to students; and c) security, which includes both reliability and validity. Additionally, Coombe *et al.* (2007) consider practicality, washback, and authenticity as well. However, despite the divisions and considerations by some authors regarding the tenets of assessment, the following five principles: validity, authenticity, practicality, washback and reliability proposed by Brown (2004) are the most used and recognized within the teaching area. Therefore, the following sub-sections describe each of them further.

#### **1.4.1. Validity**

According to Brown (2004), Coombe *et al.* (2007), Jeremy Harmer (2007), Stephen Krashen (2009), Robert Lado (1961), Richard-Amato (2005), and Ur (1999) a test is valid when it tests what it is supposed to test. This means that a test must include the contents which have been taught during the language course, using similar contexts, strategies and approaches. As Lado (1961) remarks, “validity in language tests depends on the linguistic content of the test and on the situation or technique used to test this content” (p. 30). In this light, Brown (2004) specifies that there are some requirements for a test to be valid such as having: a) a consistent structure; b) familiar formats and tasks related to their course work; c) items which are clear, simple and possible to be answered within time constraints; d) clear directions; and e) a level of difficulty with a reasonable challenge (p. 27).

Furthermore, as Brown (2004) and Arthur Hughes (1989) state, a valid test requires five components which must make possible guarantee its validity. The first component is **content validity**, which means a test has to cover the contents studied throughout the language course using well-known formats. As Coombe *et al.* (2007) exemplify, “the test assesses the course contents and outcomes using format familiar to the students” (p.

xxii). Thus, if teachers want to assess listening, they must have practiced this skill with the students before. The second component, **criterion-related validity** or criterion-related evidence, refers to the consistency of test results: There should be similar results when a test is applied under similar circumstances. In relation to this, Brown (2004) states, “criterion-related evidence is best demonstrated through a comparison of results of an assessment” (p. 24). The third element needed for a test to be valid is **construct validity** or construct-related evidence. As Brown (2004) and Hughes (1989) declare, it refers to the fact that a test has a component which measures a theoretical, non-observable construct or characteristic. As an example, it can be said that if students learnt contents through TPR (Total Physical Response), they should be assessed through the same approach used in the teaching process; in this case, TPR. In addition to this, Brown (2004) considers **consequential validity**, which includes accuracy of measuring criteria of the test, impact on the preparation of students for the test, and the impact and consequences of assessment procedures on testees' lives. In Brown's (2004) words, “[the] impact on the preparation of the test-takers, [the] effect on learners, and the social consequences of a test's interpretation and use” (p. 26) which is also part of the *washback* category (*ibid.*). Finally, the last aspect to be considered is **face validity** is related to the testees' opinions and perceptions about tests. That is, whether or not the test seems valid for the learners, if it perceives as an appropriate instrument of testing and whether or not the test items include for example, the contents studied, familiar tasks and clear instructions (Brown, 2004.).

#### 1.4.2. Authenticity

Authenticity considers the relationship among learners, their experiences and real-world language tasks (Brown, 2004). As Lyle F. Bachman and Adrian S. Palmer (1996) state, authenticity is “the degree of correspondence of the characteristics of a given



English Teachers' Assessment Practices in Valdivia in the light of the Principle of Reliability.

language test task to the feature of a target language" (p. 23). That is to say, the relationship between the test and learners' experiences are linked one with another. Consequently, teachers and test developers must contextualize the test with the students' reality in order to obtain trustworthy results. In this respect, Brown (2004) grants that an authentic language test follows five considerations: The language in the test is as natural as possible; items are contextualized rather than isolated; topics are meaningful (relevant, interesting) for the learner; some thematic organization of items is provided, and tasks that represent, or closely approximate, real-world tasks (p. 28).

As it can be seen, assessment must consider the test takers' experiences and expectations in order to be fairly assessed. By following this principle, educators assess the learners' ability to use language in context within their experience. As Lado (1961) exemplifies, "it is easy to find situations in which a person who speaks and understands a language natively fails to understand what is said because the subject matter is not within his experience" (p. 26). In summary, EFL students can perform badly in a test just because the topic of the test is beyond their experience, as well as, English speakers can performance badly for the same reason. It is important to consider real world aspect at the time of designing tests to gather consistent results.

### **1.4.3. Practicality**

Practicality plays a crucial role in language assessment due to the fact that, language tests need to be practical in relation to time, resources and in the general testing process (Brown, 2004; Coombe *et al.*, 2007; Ur, 1999). In this regard, Brown (2004) states, a practical test is seen as a tool that is inexpensive, doable within the allotted time, easy to apply and administer, and is not time consuming. That is to say, to provide a practical

English Teachers' Assessment Practices in Valdivia in the light of the Principle of Reliability.

language test, the testing procedure in terms of time and resources has to be efficient (Bachman & Palmer, 1996).

Test developers and classroom teachers need to consider time constraints as a fundamental part of the testing procedure. As Harmer (2007) indicates, it is necessary to consider "how much time should be given for test taking [ . . . ] and how much time there is for marking" (p. 386). To consider, "a test that takes a few minutes for a student to take and several hours for an examiner to evaluate is impractical for most classroom situations" (Brown, 2004, p. 19). Finally, a test is practical when teachers consider rubrics with clear answers, the time to complete a test, the time needed to score, and economic resources to design a test.

#### **1.4.4. Backwash**

Backwash, impact or washback is the influence of tests in the teaching and learning cycle (Hughes, 1989). Regarding this, Brown (2004) agrees, "washback generally refers to the effects [that] tests have on instruction in terms of how students prepare for the test" (p. 28). That is, educators must teach the students contents and test formats accordingly. Teachers must prepare test-takers before the test.

As any other principle of assessment, it can be negative or beneficial. As a negative influence educators could teach for the test contents rather than for effective learning. That is, "if a test is regarded as important, then preparation for it can come to dominate all teaching and learning activities" (Hughes, 1989, p. 1). In the same line, Harmer (2007) points out, "exam teachers suffering from the washback effect might stick rigidly to exam-format activities. In this context, the format of the exam is determining the format of the lessons" (p. 389). According to this, teaching should be based on the planning of the course, and not only on the planning of the tests. Additionally, backwash can have a

English Teachers' Assessment Practices in Valdivia in the light of the Principle of Reliability.

negative effect when students do not have any form of feedback during the preparation for the test or at the end of this process (Coombe *et al.*, 2007). Contrarily, washback has favorable impact on students when educators give constructive criticism prior and after the test. Hence, students can learn at the same time as preparing efficiently for the test.

Finally, backwash can increase “a number of basic principles of language acquisition; intrinsic motivation, autonomy, self-confidence, language ego, interlanguage and strategic investment, among others” (Brown, 2004, p. 29). These basic principles are met when constructive criticism helps educatees to focus on the aspects to be reinforced. Subsequently, test-takers apply study strategies to cope with difficult contents and to reinforce them. So, students have positive results which enhance a positive attitude towards English language (Butt, 2010).

#### **1.4.5. Reliability**

A reliable test is a language tool which is dependable and has consistent results (Brown, 2004). As Lado (1961) expresses, “reliability is measured by a correlation between the scores of the same set of students on two consecutive administrations of the test” (p. 31). As Brown (2004), Coombe *et al.* (2007), Krashen (2009), Lado (1961), and Manitoba Education, Citizenship and Youth [MECY] (2006) agree, even when testees take a test twice, the scoring should be similar. That is, the results should be similar when the same group of students takes a test twice. Additionally, Brown (2004) considers four aspects of reliability at the time of designing tests: Student-related reliability, rater reliability, test administration reliability, and test reliability.

##### **1.4.5.1. Student –related Reliability**

Student-related reliability or fluctuations in the learner refer to the testees' physical and psychological factors which influence their attitude towards the test (Brown, 2004). As

English Teachers' Assessment Practices in Valdivia in the light of the Principle of Reliability.

Krashen (2009) asserts, most students change their behavior when they know they will have a test. In the same line, Coombe *et al.* (2007) describe this as “the influences such as fatigue, sickness, emotional problems, and practice effect [which] may cause the learner’s score to deviate from the score that reflects his or her actual ability” (p. xxiii). According to Coombe *et al.* (2007), the psychological and physical aspects within the learners affect educatees’ test performance. Finally, in relation to practice effect, students’ scores can improve because they are familiar with the test format or contents, or can diminish if they are not (*ibid*).

#### **1.4.5.2. Rater Reliability**

As Brown (2004) and Coombe *et al.* (2007) agree, human error, subjectivity and bias may enter into the scoring process. Rater reliability considers human factor in the process of scoring a test. As Gary Borich and Tom Kubiszyn (1999) point out, during the preparation of the answer keys errors can be introduced. In this light, Harmer (2007) acknowledges, at the moment of scoring, bias must be out of the process. As teachers are significant elements in the scoring process, “reliability applies to both the test and the teacher” (Brown, 2004, p. 31). In the same line, Coombe *et al.* (2007), Hughes (1989) and Ur (1999) agree on the use of a variety of testing techniques which assure the best performance for both teachers and educatees. Those techniques are preparing an unambiguous rubric; take time to score; do not allow opinions of students to interfere with the scoring procedure, avoid being tired when scoring (Brown, 2004). Additionally, Brown (2004) and Coombe *et al.* (2007) divide rater reliability into two subcategories: inter-rater and intra-rater reliability.

#### **1.4.5.2.1. Inter-rater Reliability**

When assessing learners, standards of a test are also considered. This is called inter-rater reliability, which, according to Brown (2004) and Coombe *et al.* (2007) happens when two or more scorers obtain different test results when scoring the same test –score inconsistency –. It is important to acknowledge that when scorers do not follow the same standards, test inconsistency appears. It is for this reason that scorers should agree on the marking system, using the same criteria to assess students.

#### **1.4.5.2.2. Intra-rater Reliability**

Not only can students' emotional and physical factors influence their test performance, but also teacher's personal state, which is called intra-rater reliability (Brown, 2004). As Brown (2004) and Coombe *et al.* (2007) emphasize, intra-rater reliability occurs specifically when one language classroom teacher scores students based on unclear scoring patterns, bias towards learners, fatigue and/or negligence. Therefore, if a teacher is based on those aspects when assessing, scoring criteria will vary positively or negatively from one student to another.

- **Classroom Teacher Fatigue**

According to Candice Griffith and Sankaran Mahadevan (2006), “fatigue degrades human performance. The degree to which fatigue affects an individual can range from slight to catastrophic” (p.2). Fatigue is affected by circadian influence, sleep deprivation, and hours worked. All of these factors affect human performance (Griffith & Mahadevan, 2006). Among other symptoms of fatigue are found, “tired eyes, difficulty concentrating, sleepiness and muscle pain” (Drave, n.d, p. 2).

Research done by Helen Aslett (2006) evidences, cognitive aspects such as fatigue and lack of concentration affect the scoring process. In Aslett's (2006) words, “the process

English Teachers' Assessment Practices in Valdivia in the light of the Principle of Reliability.

of how marking accuracy of a single assessor decreases over time due to fatigue and lack of attentional control” (p. 86). In this light, the effects of fatigue on the marking procedure produce an inconsistency of the tests results, and so problems with reliability (Brown, 2004; Coombe *et al.*, 2007). In addition, studies done by June Pilcher and Allen Huffcutt (1996) prove that sleep deprivation contributes to fatigue and has negative influence on task performance (as cited in Aslett, 2006, p. 87). With this regard, the American Psychological Association [APA] (2012) states, “sleep is essential for a person’s health and wellbeing. Yet millions of people do not get enough sleep and many suffer from lack of sleep” (para. 1).

In order to avoid fatigue during the grading procedure, Coombe *et al.* (2007) empathizes, teachers need to “make sure [they] can locate [the answer key] quickly so that marking can commence in a timely manner” (p. 153). Furthermore, it can be stated that teachers need to sleep appropriately, because fatigue is measurable as a reaction time factor on human performance (APA, 2012; Aslett, 2006; Griffith & Mahadevan, 2006). So, in order to obtain reliable test results, tests must be during the day.

- **Marker Negligence**

Negligence is seen as the lack of care or attention towards the scoring process (Brown, 2004; Coombe *et al.*, 2007). The cause of this careless behavior, it is exemplified by Lado (1961) “the examiner hesitates..., wonders if this response is as good as another he considered good, if he is being too easy or too harsh in his scoring” (p. 31). As it infers, negligent behavior appears when tests do not have an answer key, rubrics or any other pattern of scoring in order to assure reliability of the test results. As Brown (2004) emphasizes, a scoring instrument is needed to assure intra-rater reliability. Additionally,

English Teachers' Assessment Practices in Valdivia in the light of the Principle of Reliability.

Coombe *et al.* (2007) recommend, “after the test has been administered, teachers must mark the papers using the answer key prepared at the same time the test was develop” (p. 153).

Negligent behavior can be avoided when classroom teachers design, apply and grade their own tests (Brown, 2004). However, this process has to be as objective as possible. As Lado (1961) mentions, “objective tests are easy to score” (p. 31), because at the moment of grading tests, teachers do not spent time wondering if they have been “too easy” or “too harsh”, because they use a rubric to avoid subjectivity (Brown, 2004; Lado, 1961).

- **Bias towards Students**

Preferences towards a particular pupil based on personal opinions are called bias. Bias can have a positive or negative influence for students, which interferes on the teachers' assessment practices (Brown, 2004; Coombe *et al.*, 2007; Popham, 2012). General factors that promote bias are “culture, developmental stage, ethnicity, gender, socioeconomic background, language, special interests, and special needs” (MECY, 2006, p. 9). However, in James Popham's (2012) words three common assessment biases appear in educational settings: “Racial/ethnic bias, gender bias, and socioeconomic bias” (p. 5). This psychological phenomenon affects assessment practices and students tests results (Aslett, 2006). “The teachers' unconscious gender biases can produce stereo-typic expectations for students' success and participation in the classroom [*sic*]” (Scantlebury, 2009, para. 5). Cultural or racial/ethnic bias is “the belief that the dominant or mainstream (presumably European and North American) cultural ways of learning and knowing are superior” (Stevens, Tyler & Uqdah, 2009, para.1). Finally, socio economic bias is evidenced by “indicators such as education, wealth, inheritance, and occupation” (Nova Scotia Department of Education [NSPE], 2001, p. 15).

In terms of the test itself, Popham (2012) identifies that bias can be found when the test has offensive items, and unfair penalization. He refers to unfair penalization as the use of items that impedes one or more subgroups of students from performing adequately. In relation to this, Aslett (2006) mentions, during the grading process “where a marker is aware of a student’s identity, their marking can potentially be profoundly affected” (p. 87). Another component that influences the intra-rater reliability of tests is the experience of the marker. Kathryn Ecclestone (2001) points out, “novice markers are more accurate compared to experienced markers who placed greater importance on their intuition” (as cited in Aslett, 2006, p. 88). Hence, if markers rely on their educational experience rather than on objective criteria the reliability of tests is not met.

In order to banish bias, it is recommended to have a scoring instrument (Aslett, 2006; Coombe *et al.*, 2007; Richard-Amato, 2005). To not include offensive items, contents which insult a particular group of students, vocabulary related to gender, and any items that could have a negative impact on test-takers (Coombe *et al.*, 2007; Popham, 2012). Furthermore, it can be necessary not to look at the students’ names at the time of grading tests (Aslett, 2006).

#### **1.4.5.3. Test Administration Reliability**

Apart from student-related and rater-reliability, environmental factors must be considered as part of the process of assessing (Brown, 2004). In this sense, test administration reliability happens when a test is taken under circumstances which could affect testees’ concentration on the test; for example, desk arrangement, poor light condition, street noise and drastic variations of temperature (Brown, 2004; Coombe *et al.*, 2007). According to Coombe *et al.* (2007), fluctuations in the test administration occur also when “different students are tested in different locations on different days” (p. xiv).



- **Classroom Lighting Condition**

Appropriate lighting quality is defined as “visual, comfort, good color, uniformity and balanced brightness” (Northeast Energy Efficiency Partnership [NEEP], 2002, p.1). Studies done by the Philips Electronics Company [PEC] (2010) indicate, classrooms having high-quality lighting system show improvements on students' concentration, and behavior. Also, it mentions that when pupils have better light conditions – both natural and artificial – they perform obtain better learning results.

In Brown's (2004) words, unreliability is found when students answer tests with variations of light which affects students' performance. In this regard, the Osram Sylvania Company [OSC] (2012) agrees, lighting needs to be equally on each student's work area. For this reason, at the moment of applying tests, educatees need suitable light conditions to enhance their testing performance (Coombe *et al.*, 2007). In this regard, the National Clearinghouse for Education (NCEF) (n.d) points out, “there is a ... proven link between illumination and student achievement” (as cited in McCreery & Hill, para. 2).

Favorable lighting conditions promote learning (Brown, 2004; NEEP, 2002; PEC, 2010). It can be stated that “lighting plays a positive part in that, keeping [students] young minds alert and eager to take part” (PEC, 2010, p. 3). Modifications on behavior and performance demonstrate that luminance, both natural and artificial, plays a relevant role in educational and psychological matters (Brown, 2004; Coombe *et al.*, 2007).

- **Classroom Intrusive Noise**

Noise impacts negatively on test administration reliability of test results (Brown, 2004; Coombe *et al.*, 2007). As Brown (2004) exemplifies, if during a listening test, test-takers are close to the windows, they will listen to the street noise and, so they will lose concentration. This fact will interfere with educatees English language performance.

English Teachers' Assessment Practices in Valdivia in the light of the Principle of Reliability.

Supporting this, Julie Dockrell and Bridge Shield (2008 a) state, “environmental and classroom noise have detrimental effects upon children’s academic performance; and also that noise has more of an impact upon children with special educational needs than upon other children” (para. 2). Additionally, external noise has higher negative effect on older pupils (Dockrell & Shield, 2008 b; Partin, 2009).

To impede students distraction at the moment of taking tests, “the test room should be free from interruptions and background noise” (Coombe *et al.*, 2007, p. 149). In this line, to reduce external noise some building modifications can be made; for example, “the use of acoustic ceiling tiles (Maxwell and Evans, 2000), wall coverings and carpets to absorb sound (Tanner and Langford, 2002)” (as cited in Alexander, 2010, p. 599).

However, classroom teachers have no implication on this matter, since noise control consultants and architects must create a suitable learning environment for students to have an adequate learning environment (Alexander, 2010).

- **Classroom Seating Arrangements**

Desks are used to accomplish a testing environment at the moment of applying tests (Brown, 2004; Coombe *et al.*, 2007; Partin, 2009). As Coombe *et al.* (2007) state, “the setting of a test should provide an atmosphere that is conducive to student learning” (p. 149). Behavior, learning and group dynamics of the students are influenced by the physical desk arrangement of a classroom (Denton, 1999; Partin, 2009).

To promote reliability of test conditions, “the ideal seating is every other seat and every other row” (Coombe *et al.*, 2007, p. 149). In this light, Ronald Partin (2009) mentions, the greater distance between students desks at the time of taking tests facilitate the overall assessment process. Additionally, he emphasizes that rows minimizes distractions, so it produces higher levels of on-tasks behavior, and affect positively students

English Teachers' Assessment Practices in Valdivia in the light of the Principle of Reliability.

with special needs (Partin, 2009). Coombe *et al.* (2007) adds, adequate desk arrangement prevent cheating at the moment of answering tests. For these reasons, “the arrangement of students’ desks should not be permanent, except for large lectures halls or laboratories” (Partin, 2009, p. 3). Therefore, the nature of the tasks should dictate the most convenient seating arrangement for students to enhance their learning experience (Ruhl & Wannarka, 2008).

Classroom control is encouraged by desk arrangement during tests (Black, Lambert, & Rosenfield, 1985; Denton, 1992, Coombe *et al.*, 2007). The physical desk arrangement of classrooms has “significant implications for discipline and classroom management” (Denton, 1992, p. 29). Thus, teachers must be aware of desk arrangement at the moment of taking tests since students recognize and are influence by the testing environment of a classroom (Brown, 2004; Coombe *et al.*, 2007; Partin, 2009; Ruhl & Wannarka, 2008).

#### **1.4.5.4. Test Reliability**

The last, but not least, aspect needed for a test to be reliable corresponds to test reliability. This sub-principle considers the conditions of a test itself such as item design, test length and text format. Firstly, it can be said that if there is more than one correct answer, the test items have been poorly designed. In addition, when a test is too long for students to finish, it can be considered as an unreliable assessment tool. Finally, it is also essential to consider the readability of the text, as small fonts or blurred text may impact negatively on learners’ performance (Brown, 2004).

As a conclusion, the use of well-designed tests and the awareness of the principles of assessment help educators gather objective information about educatees’ learning processes and test design (Richard-Amato, 2005). The importance of well-designed tests resides in the impact of tests on test-takers’ lives, “... the huge impact that assessment outcomes have

English Teachers' Assessment Practices in Valdivia in the light of the Principle of Reliability.

on an individual's learning career, their motivation to engage with educational processes and on their opportunities in life" (Butt, 2010, p. 121). For this reason, teachers need to demonstrate that results are reliable (Bachman, 2004, p. 3).

## **2. CHAPTER II: Methodology**

### **2.1. Justification**

#### **2.1.1. Assessment in Latin America**

“Tests are essential components of a successful curriculum and one of several partners in the learning process” (Brown, 2004, p. 16), not only because they measure students’ academic performance, but also because they reflect the reality of a country. In Guillermo Ferrer’s (2006) words, “the models and systems designed for measuring and assessing student academic performance obviously reflect a country’s political conditions and technical and operational capacities, both current and potential” (p. 9).

In order to obtain results, tests have been used as the first instrument to assess all kinds of knowledge, as for example, EFL students’ knowledge about English. However, little importance has been given to assessment practices in Latin America, in general, because “insufficient thought has been given to the specific purposes assessment systems are expected or desired to serve” (Ferrer, 2006, p. 15), among other elements involved. In contrast, Chile as well as Colombia have been pioneers in terms of assessment, “[they] have been continuously administering tests of academic achievement for many years, and their assessment agencies enjoy a high degree of legitimacy and continuity despite the political vagaries periodically prompted by changes in national government [*sic*]” (Ferrer, 2006, p. 16).

#### **2.1.2. English Assessment in Chile**

In terms of EFL testing, Chile has only applied two English proficiency tests, whose results were poor (EducarChile, 2011). Although the MINEDUC has implemented some programs to improve the English teaching praxis in Chile, little attention has been paid to assessment, as an essential factor to improve students’ level of the English language. This

English Teachers' Assessment Practices in Valdivia in the light of the Principle of Reliability.

means that the national curriculum does not address language accurately: Apart from focusing on receptive skills –and “any test which assesses only reading and listening cannot really be considered an adequate test of proficiency” (Sheehan, 2011). In Sheehan’s (2011) words, there is a lack of assessment expertise. This factor is worrying if we think that the MINEDUC administers and regulates assessment in Chile (A. Sheehan, personal communication, June 21, 2011).

## **2.2. Educational Survey Research**

This educational survey research (Gable, 1994; Kraemer, 1991; UNESCO, 2005) collected information through educational surveys that sought to find out about teachers’ assessment practices in the Commune of Valdivia, especially focused on the principle of reliability (Brown, 2004; Coombe *et al.*, 2007; Ur, 1999). Results were quantitatively analyzed and shown in pie charts. Additionally, comments on the surveys were addressed qualitatively since those are based on teachers’ empirical basis. Kenneth Kraemer (1991) identifies three distinguishing characteristics of survey research. Firstly, it is employed to quantitatively describe specific aspects of a given population. Secondly, the information looked for through the survey is collected from people and is, therefore, subjective. Finally, it is believed that the subject sample’s ideas can be generalized.

The survey method considers the design of questions that support the research. In this particular case, the questions were intended to address the two sub-principles of assessment: intra-rater and test administration reliability posed by Brown (2004) (See Appendix II).

### **2.3. Data Collection Method**

This research was conducted in order to determine if the principle of assessment, reliability –intra-rater and test administration reliability especially– proposed by Brown (2004) is met at the time of assessing students of the English language in Valdivia.

A survey was applied to 26 English teachers in the Commune of Valdivia. This single procedure consisted of 14 close-ended questions and one comment section (See Appendix II) related to the principles of assessment in general; however, this study only focuses on intra-rater and test administration reliability, so only questions addressing those sub-principles are analyzed and discussed later.

The survey has two sections: Section I, which is intended to identify the educator; and Section II, which is focused on intra-rater and test reliability (Brown, 2004). At first, this tool was applied using electronic mail, and as second instance, samples were directly administered to English teachers from different Valdivian educational settings. Closed-ended questions were chosen because they are less time consuming, more effective and more economically convenient (Bailey, 1994).

### **2.4. Data Collection Analysis**

Data was analyzed from a quantitative point of view. Results were summarized and tabulated (UNESCO, 2005); answers were classified into positive and negative (yes or no) and/or options, which were used as indicators, which refer “to one or more pieces of numerical information related to an entity that one wishes to measure” (UNESCO, 2005, p. 42).

Finally, generalizations of the EFL teachers were made, in order to describe the general profile of English teachers in Valdivia. Hence, there is a high probability of drawing

English Teachers' Assessment Practices in Valdivia in the light of the Principle of Reliability.

generalizations about the target population in terms of assessment practices by using descriptive analyses.

## 2.5. Subjects

The subjects of this survey research were 26 English Foreign Language teachers from the Commune of Valdivia. These key participants were mainly teachers from Valdivia, Futrono, San Jose de la Mariquina and Paillaco. The answers of 11 English teachers were received by electronic mail in which three surveys were answered throughout the “English across the Rivers” Network Coordinator, Valdivia. Then, the surveys were directly administered to 15 English teachers from 10 different Valdivian educational settings. Teachers taught from pre-kinder to 12<sup>th</sup> grade.

Finally, the survey involved 19 schools (See Appendix I): 8 public and 11 semi-private educational settings, in which 11 teachers work in public and 15 teachers in subsidized schools. Bellow you will find a table with summarized information.

Table 1

### *English Teachers Profile*

<b>Educational settings</b>	<b>Public or Subsidized</b>	<b>Amount of Students (approx.)</b>	<b>Grades</b>
Escuela España	Public	39	Pre-kinder, 1 <sup>st</sup> , 2 <sup>nd</sup> , 3 <sup>rd</sup> , 4 <sup>th</sup>
Escuela España	Public	35	5 <sup>th</sup> , 6 <sup>th</sup> , 7 <sup>th</sup> , 8 <sup>th</sup>
Instituto Salesiano	Subsidized	32	1 <sup>st</sup> , 2 <sup>nd</sup> , 3 <sup>rd</sup> , 4 <sup>th</sup> , 5 <sup>th</sup> , 6 <sup>th</sup> , 7 <sup>th</sup> , 8 <sup>th</sup>
Colegio Laico	Subsidized	24	2 <sup>nd</sup> , 3 <sup>rd</sup> , 4 <sup>th</sup> , 5 <sup>th</sup> , 6 <sup>th</sup>
Colegio Principe de Asturias	Subsidized	25	1 <sup>st</sup> , 2 <sup>nd</sup> , 3 <sup>rd</sup> , 4 <sup>th</sup>
Colegio San Lucas*	Subsidized	22	Pre-Kinder, 1 <sup>st</sup> , 2 <sup>nd</sup> , 3 <sup>rd</sup> , 4 <sup>th</sup> , 5 <sup>th</sup> , 6 <sup>th</sup>
Colegio Particular	Subsidized	31	1 <sup>st</sup> , 2 <sup>nd</sup> , 3 <sup>rd</sup> , 5 <sup>th</sup>



English Teachers' Assessment Practices in Valdivia in the light of the Principle of Reliability.

Gracia y Paz			
Instituto Tecnológico del Sur	Subsidized	18	5 <sup>th</sup> ,6 <sup>th</sup> ,7 <sup>th</sup> ,8 <sup>th</sup>
Escuela Proyecto Futuro	Public	38	Pre-Kinder, 1 <sup>st</sup> ,2 <sup>nd</sup> ,3 <sup>rd</sup> ,4 <sup>th</sup>
Liceo Santa María La Blanca	Public	28	10 <sup>th</sup> ,11 <sup>th</sup>
Instituto Superior de Administración y Turismo	Public	32	9 <sup>th</sup> ,11 <sup>th</sup> ,12 <sup>th</sup>
Instituto Superior de Administración y Turismo	Public	28	9 <sup>th</sup> ,10 <sup>th</sup> ,11 <sup>th</sup> ,12 <sup>th</sup>
Instituto Italia	Public	37	9 <sup>th</sup> ,10 <sup>th</sup> ,11 <sup>th</sup> ,12 <sup>th</sup>
Instituto Comercial	Public	40	9 <sup>th</sup> ,10 <sup>th</sup> ,11 <sup>th</sup> ,12 <sup>th</sup>
Liceo Polivalente Los Avellanos	Public	21	9 <sup>th</sup> ,10 <sup>th</sup> ,11 <sup>th</sup> ,12 <sup>th</sup>
Colegio Helvecia	Subsidized	28	5 <sup>th</sup> ,6 <sup>th</sup> ,7 <sup>th</sup> ,8 <sup>th</sup> 9 <sup>th</sup> ,10 <sup>th</sup> ,11 <sup>th</sup> ,12 <sup>th</sup>
Colegio Helvecia	Subsidized	24	3 <sup>rd</sup> ,6 <sup>th</sup> ,7 <sup>th</sup> ,8 <sup>th</sup> 9 <sup>th</sup> ,10 <sup>th</sup> ,11 <sup>th</sup> ,12 <sup>th</sup>
Colegio Austral	Subsidized	25	7 <sup>th</sup> ,8 <sup>th</sup> 9 <sup>th</sup> ,10 <sup>th</sup> ,11 <sup>th</sup> ,12 <sup>th</sup>
Liceo Santa María La Blanca	Public	20	7 <sup>th</sup> ,8 <sup>th</sup> 9 <sup>th</sup> ,10 <sup>th</sup> ,11 <sup>th</sup> ,12 <sup>th</sup>
Colegio Alonso de Ercilla*	Subsidized	37	7 <sup>th</sup> ,8 <sup>th</sup> 9 <sup>th</sup> ,10 <sup>th</sup> ,11 <sup>th</sup> ,12 <sup>th</sup>
Colegio José Manuel Balmaceda*	Public	30	4 <sup>th</sup> ,5 <sup>th</sup> ,6 <sup>th</sup> 9 <sup>th</sup> ,10 <sup>th</sup> ,11 <sup>th</sup> ,12 <sup>th</sup>

English Teachers' Assessment Practices in Valdivia in the light of the Principle of Reliability.

Colegio María Auxiliadora	Subsidized	35	1 <sup>th</sup> , 2 <sup>nd</sup> 9 <sup>th</sup>
Colegio María Auxiliadora	Subsidized	37	3 <sup>rd</sup> , 4 <sup>th</sup> , 5 <sup>th</sup> , 6 <sup>th</sup> 9 <sup>th</sup> , 10 <sup>th</sup> , 11 <sup>th</sup>
Liceo Técnico Profesional Mater Populi Dei	Subsidized	30	9 <sup>th</sup> , 10 <sup>th</sup>
Instituto Tecnológico del Sur	Subsidized	27	9 <sup>th</sup> , 10 <sup>th</sup> , 11 <sup>th</sup> , 12 <sup>th</sup>
Instituto Salesiano	Subsidized	40	3 <sup>rd</sup> , 4 <sup>th</sup> , 5 <sup>th</sup> , 7 <sup>th</sup> 9 <sup>th</sup> , 10 <sup>th</sup> , 11 <sup>th</sup> , 12 <sup>th</sup> ,

*Note:* Teachers giving English lessons in both cycles were considered as part of secondary system (primary and secondary education).

\*Surveys answered by e-mail throughout the “English across the Rivers” Network Coordinator.

### **3. CHAPTER III: Results**

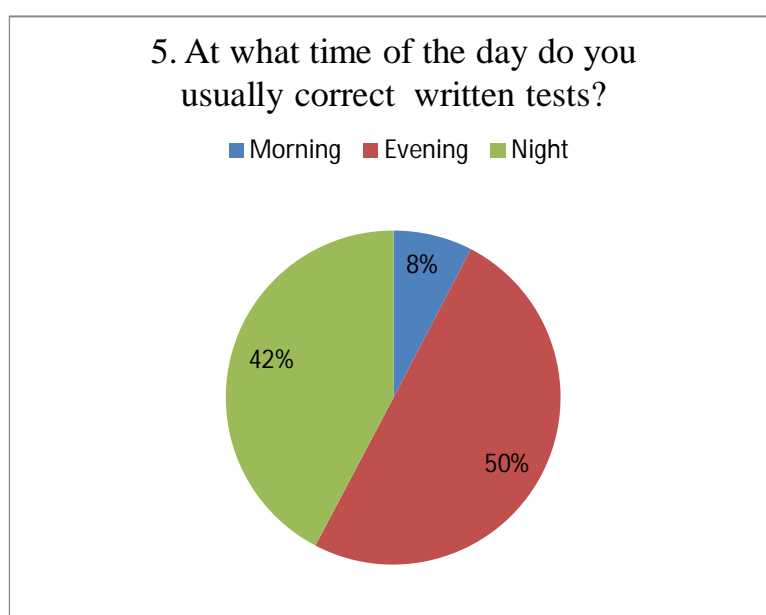
This research study looked for depicting EFL teachers' general assessment practices when giving written tests (commonly known as pen-and-pencil tests) in the Commune of Valdivia, especially in the light of the intra-rater and test administration reliability concepts developed by Brown, 2004.

Section I in the survey helped create a profile of teachers and classroom conditions (amount of students and grades) in Valdivia. This information was tabulated (See Table I). In summary, 17 high school teachers – eight EFL teachers from public and nine from subsidized high schools – answered the survey. The average amount of students per class was 30 students in the public high schools and 31 students per class in subsidized high schools. In relation to primary school, nine school teachers – three English teachers from public and six from semi-private schools – participated in this research, showing that the average number of students was 31 in public and 25 students in subsidized primary schools. In general, English educators in the commune of Valdivia have 32 students per class in public educational settings and 29 students per class in semi-private education. Regarding the amount of courses per teacher, teachers have four courses in public and six courses in semi-private high schools, while in primary the amount of courses is five courses in public and five courses in subsidized educational settings. In total, secondary teachers have five courses and primary teachers, five courses.

Regarding Section II, only questions **5, 6,7,11, 12, and 14** were analyzed, since they directly address the subject matter of this paper. Results from this section are represented below by means of pie charts.

Questions 5, 6 and 7 give information about the sub-element **intra-rater reliability** of written tests. Aspects such as fatigue, negligence, and bias towards students were respectively analyzed.

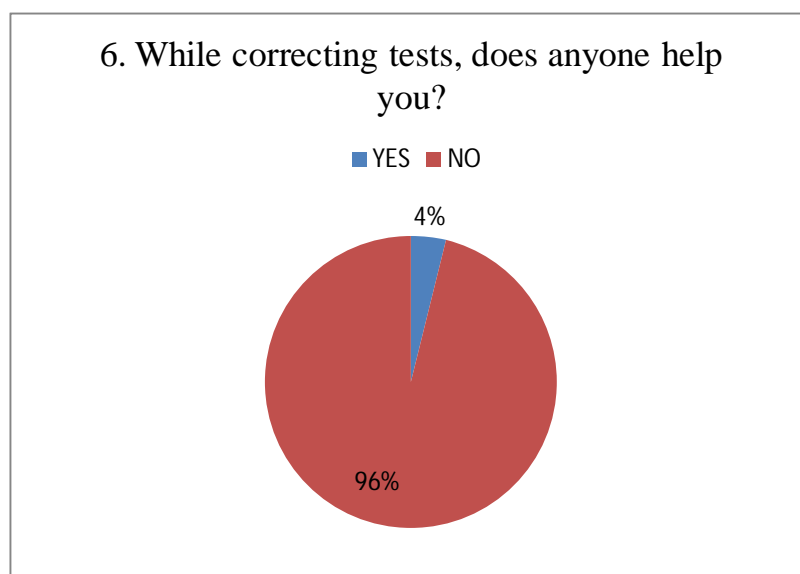
Question 5 addressed intra-rater reliability; it specifically looked for knowing if teachers were tired or not at the moment of grading tests.



Graph 1: Comparison among the English teachers marking preferences

Question 5 revealed that 50% of the subjects check tests in the evening, 42% at night and only 8% in the morning, which might lead to “harder” grading criteria, in Brown’s (2004) words, also revealing that the sub-principle of intra-rater reliability might be affected by teachers’ assessment practices.

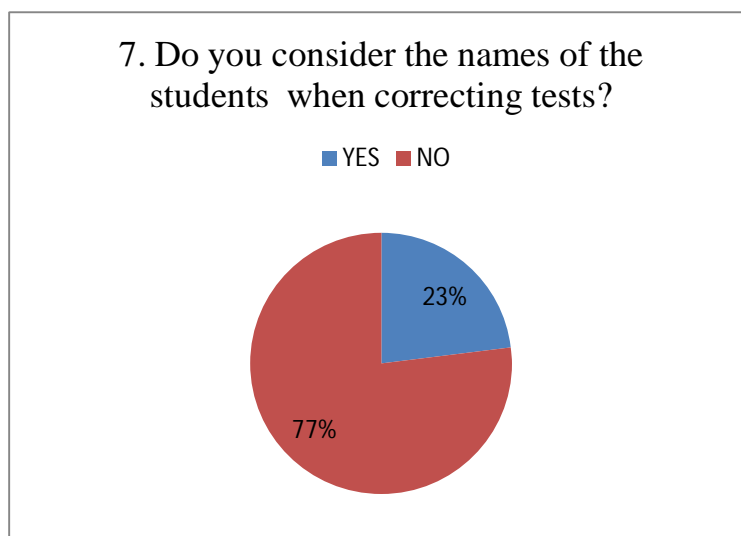
Question 6 was focused on one aspect of intra-rater reliability (negligence) at the moment of correcting written tests. This question gathered information to find out if there was more than one scorer at the time of checking tests.



Graph 2: English teachers in relation to extra help into the scoring process

Outcomes showed that 96% of the English teachers check tests personally. Only 4% of teachers were helped by teaching practice students. However, those English students design, apply and grade the tests, so there is no negligence at the moment of assessing students.

The last aspect of intra-rater reliability was focused on whether or not EFL educators are biased against students when assessing their results, which are summarized in Graph 3. In particular, teachers were asked if their students' names affected their grading process in some way.



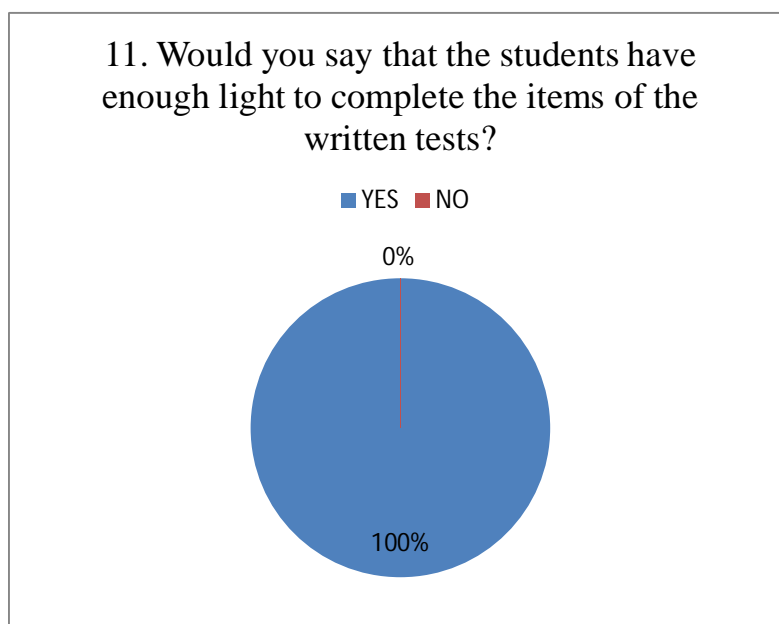
Graph 3: Influence of the name of educatees on the scoring process

Findings evidenced that 77% of the subjects did not look at the names of the students in the tests. However, 23 % of the teachers did consider students' names. In Coombe *et al.*'s (2007) words, if the scorer is prejudiced against students, it could affect tests results and so reliability.

To conclude, the sub-element intra-rater reliability was portrayed through three pie charts showing that 50% of the subjects graded tests during the evening, 96% did not require help during the marking process, and 77% did not consider the name of the students at the time of checking the tests.

The second part of this research was focused on **test administration reliability**. Questions 11, 12 and 14 are also represented through charts. Aspects such as light condition, noise and seating arrangement were analyzed.

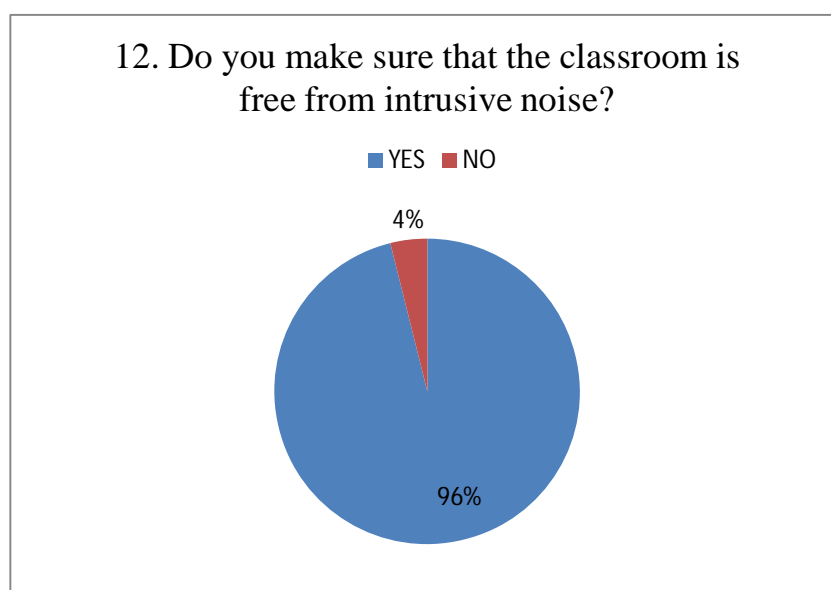
Question 11 collected information about light conditions in the English classroom at the moment of applying written tests. Graph 4 summarizes results.



Graph 4: This chart shows that students have enough light to answer tests

Results showed that 100 % of teachers considered light conditions at the time of applying tests. As it can be seen, the use of adequate light conditions is a substantial element for teachers at the moment of giving tests (Northeast Energy Efficiency Partnership, 2002).

Question 12 gathered information about the presence or absence of intrusive noise when students completed tests. Graph 5 summarizes this information.

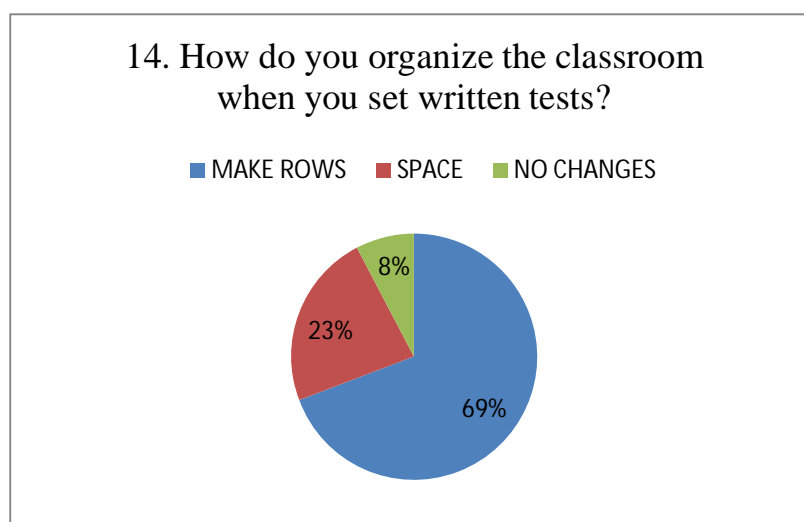


Graph 5: Intrusive noise during the test taking procedure

Findings demonstrated that 96% educators maintained the English classroom without noise at the time of giving written tests, while only 4 % did not consider this factor. This 4% of answers corresponded to primary school teachers. It could be said that EFL teachers consider an appropriate testing atmosphere, then.



The last element to be considered as part of test administration reliability was whether or not English teachers made use of testing seating arrangements (Question 14).



Graph 6: Classroom arrangement at the time of taking tests.

Outcomes illustrated that 69% of educators organized students in rows, 23% left a space between test-takers and 8% did not change the classroom for test procedures. The 8% of teachers corresponded to primary school teachers while the rest of the teachers made changes at the moment of setting written tests. However, the percentages showed that the majority of the Valdivian educators considered testing seating conditions.

To sum up, the **test administration reliability** results demonstrated that 100% of the educators had classrooms with appropriate light conditions, 96% of the teachers kept the classroom without noise that could interfere with the students' test taking procedure, and 69% of the teachers organized their students in rows during tests.

#### **4. CHAPTER IV: Discussion**

As it was already mentioned, the principle of reliability is composed of four elements: Students-related reliability, rater reliability, test administration reliability and test reliability (Brown, 2004). However, the aim of this study is to find out whether or not English language teachers meet two aspects of the principle of reliability: Intra-rater and test administration reliability at the time of assessing their students.

The information based on the results yielded by the surveys applied to 26 English teachers will be discussed in order to diagnose the general assessment practices in Valdivia, to generalize it to the Chilean context.

##### **4.1. Intra-rater Reliability**

Through the analysis of the surveys, aspects of intra-rater reliability, such as fatigue, negligence and bias were analyzed. Results revealed that teachers are highly prone to checking tests in the evening and at night (92% in total), which could lead them to suffering from fatigue or tiredness (Aslett, 2006; Brown, 2004). In terms of intra-rater reliability, fatigue is one factor that promotes unclear scoring patterns due to lack of motivation and energy (Brown, 2004; Coombe *et al.*, 2007). In this regard, tests that are scored first could have different criteria than the ones scored later (Lado, 1961). As Brown (2004) exemplifies, “I may be “easier” or “harder” on those first few papers or I may get tired, and the results may be an inconsistent evaluation across all the tests” (p. 21). Following this, errors in measuring due to fatigue especially during evenings and at bedtime do not assure intra-rater reliability in tests. “Mental fatigue due to monotony and lack of interest in a task can have severe implications with regards to task performance and accuracy” (Aslett, 2006, p. 86). Apparently, Valdivian teachers do not have time to grade tests during the working

period due to a heavy academic burden. In informal conversations with some of the teachers involved in this study, they commented that they needed to have time for meeting parents, designing tests, calling the roll, asking students to behave, and preparing material, among other tasks, but there was no time available to grade tests in the school; for that reason, tests were checked at home and later on the day (Dec., 2012).

The second aspect of intra-rater reliability to be analyzed is negligence; surveys demonstrated that classroom teachers were the only responsible people for the grading criteria most of the times (96%). However, some times, there were teaching practice students involved in the process (4%), although the principle of intra-rater reliability was also met, because they designed, applied and checked the tests themselves (Brown, 2004). Comments in informal conversations support this: Teachers disclosed that two or more scorers were almost never involved. In this regard, let us remember that in order to achieve intra-rater reliability, only one scorer must mark the tests (Brown, 2004; Coombe *et al.*, 2007). In this manner, it can be said that teachers accomplish this sub-element of intra-rater reliability.

Another factor involved in the study was whether or not teachers have preferences regarding students; in other words, bias. The survey showed that bias towards “good” or “bad” students was not a common praxis among teachers; in fact, 77% said they did not look at their students' names. Teachers commented in the surveys that tests were checked by item and then graded; it was not time efficient to look at each pupil's names. Others said that it was a poor pedagogical practice to consider the names of the students. However, under specific circumstances the names of students were considered (23%). For instance, two teachers pointed out that it was inevitable for them to see the names of the students in a

English Teachers' Assessment Practices in Valdivia in the light of the Principle of Reliability.

test. Others said that they wanted to follow students' progress or because they worked with special needs students. Finally, teachers working with students in governmental integration programs agreed that they had to see the students' names, because the educational setting asked them to track students' progress.

In summary, results yielded that the principle of **intra-rater reliability** was met up to a certain point. In terms of fatigue, it was shown that teachers assessed students at an inconvenient time, being evening or night, which may lead to unfairness. However, it was also found out that this practice was due to lack of time, in general. Teachers did not find time during their working hours to correct tests, so they had to bring them home. No negligence was found. In this regard, teachers were responsible enough to check their own tests and if someone else happened to check them, it was because they had designed them and applied them. Finally, it seems that bias towards students was not involved in the process of assessment, since teachers tried not to consider the students' names when correcting tests, which adds to the principle of reliability. So, in general, it can be said that the sub-principle of intra-rater reliability was actually met. However, it would be quite interesting to contrast teacher-related results to what students have to say, since on many occasions they complain about teachers' assessment practices, accusing them of being unfair.

#### **4.2. Test Administration Reliability**

The outcomes related to test administration reliability demonstrated that classroom teachers promote the appropriate use of light conditions at the time of applying written test. This is important since test-takers need high-quality illumination to answer tests correctly “to promote learning, provide an environment where teachers and students can perform their visuals tasks comfortably, quickly and accurately” ( Northeast Energy Efficiency

English Teachers' Assessment Practices in Valdivia in the light of the Principle of Reliability.

Partnership, 2002, p. 8). Comments in informal conversations (Dec., 2012) showed that having a classroom with good-quality light was something inherent to the classrooms. However, structural building factors determine the adequate or inadequate luminance of a classroom. As Mark Winterbottom and Arnold Wilkins (2009) point out, luminance problems appear due to poor policy decisions when designing educational settings. In this regard, teachers have no part of this issue in terms of structural aspects. In spite of this, EFL teachers control the artificial lighting and supply the lack of light in the classrooms and/or prevent the excess of luminance at the time of applying tests.

The lack of intrusive noise is fundamental when testing students. In this light, comments in informal conversations (Dec., 2012) showed that subjects used practical thinking at the time of setting tests. Teachers mentioned that this is an essential part of the teachers' roles, since the lack of students' concentration is reflected on the grades. As Harmer (2007) emphasizes, "moments of disruptions can be unsettling, not just for teachers but also for students" (p. 153). Additionally, some of the subjects agreed that assessment theories are applied without teachers even knowing them. That is to say, common sense is appreciated in the teaching practices of the Valdivian teachers.

Seating arrangement is also seen as an important part of the teaching field. In this regard, as Alexandra Ramsden (1999) states, "the considerations in arranging the physical environment of the room is so that teaching and learning can occur as efficiently as possible" (para. 1). In general, results showed that the seating arrangement procedure is a common practice among teachers; on the contrary, the inexistence of a testing atmosphere could lead to misunderstanding, especially in the context of primary school students, since they may not have developed an idea of what test-taking involves (Denton, 1992; Partin, 2009). Norris Minnick (1999) agrees, "for children, the word has no meaning outside the

English Teachers' Assessment Practices in Valdivia in the light of the Principle of Reliability.

concrete context in which it is used” (as cited in Lotfy, 2012, p. 2). In relation to this, it is important to mention that only 8% of the primary teachers involved in this study did not show interest in making classroom changes. This is relevant since when setting tests, the teacher needs space to walk around the classroom (Ramsden, 1999) in order to maintain discipline and assist students during the tests. Besides, arranging the classroom in a special way helps primary students understand that testing is a formal situation in which they need to behave in a particular way to obtain positive results. Testing environments promote testing behavior. According to Ashley Borders, Mitchell Earleywine, and Stanley Huey (2004), “individuals choose behaviors based on the outcomes they expect and the values they ascribe to those expected outcomes” (as cited in Cruz, 2005, p. 3).

High school teachers stated that they did as much as possible to avoid copying, lack of concentration and unpleasant behavior when testing, since those elements interfere with the test-takers performance (Dec., 2012). At this point, high school students should naturally maintain a test-like atmosphere. In relation to this, Kathy Escamilla and Elizabeth Grassi (2000) state, “it is important that the teacher maintain[s] a relaxed and enjoyable learning environment” (p. 123) so that students behave properly and show what they really know or what they can actually do.

In a nutshell, with reference to test administration reliability, it seems that this principle is generally met. Light conditions, noise, and seating arrangement were analyzed. It was found that classroom teachers sought to avoid any type of situation that could affect test-takers negatively. Quite on the contrary, it seems that teachers looked for the best conditions for their students to take tests because, as Coombe *et al.* (2007) agree, “fluctuations in test administration, inconsistent administrative procedures and testing conditions will reduce test reliability” (p. xxiv).

English Teachers' Assessment Practices in Valdivia in the light of the Principle of Reliability.

There was no evidence of lack of **test administration reliability** since the role of common sense in the Valdivian educational practices seems to be extremely important (Dec., 2012). However, some discrepancies were found among primary school teachers when giving tests; in other words, all high school teachers took all steps to meet this principle, although some primary school teachers did not.

Therefore, teachers in the Commune of Valdivia partially met the principle of intra-reliability. The main problem was found with regards to the time used to check tests, since fatigue interferes with test results. Tests must be marked during mental concentration periods, which are during light time. "Fatigue, either mental (lack of interest/repetition) or physical (lack of sleep), has been found to significantly affect the reliability of the marks assigned by an individual assessor" (Aslett, 2006, p. 86). However, in spite of this aspect of intra-rater reliability, the principle of test administration reliability was totally met.

It is essential to meet most of the principles of assessment in order to contribute to the reliability of tests results, thus making tests reliable tools to measure educatees' real performance in a fair way. The role of a reliable assessment in the teaching field is to help educators identify students' strengths and weaknesses, monitor students' progress, plan and conduct class instruction (Educational Testing Service, 2003). Apart from that, the improvement of students' performance, the increment of equity and the use of test-takers techniques are among the benefits found in relation to educatees (OECD, 2005). It is necessary to give students an appropriate context to be tested to assure the reliability of results. Finally, not only do students need appropriate test conditions but also a truthful teacher to mark tests.

To finish with this, it seems significant to refer to a particular situation that might have affected the results in this study. The primary subjects of this research were the

English Teachers' Assessment Practices in Valdivia in the light of the Principle of Reliability.

teachers participating in the English teaching network in Valdivia. Those subjects were considered pertinent to be analyzed, since it was supposed that they were interested in the assessment matters. As they participated in the English network, it could be inferred that they wanted to improve the English teaching education system as a whole. In addition, these subjects were taken into consideration due to the fact that they come from the Provinces of Valdivia and Futrono, and it would have been very useful if all of them had participated in this study, in order to have an overall view of the region.

In October 18<sup>th</sup>, 2012, the Coordinator of English across the Rivers (the abovementioned network) was contacted via electronic mail to see if the surveys could be answered in the English network meeting in November or December 2012. She offered to send the surveys by electronic mail and to encourage teachers to answer the survey via the same means. In total, throughout the coordinator, only three English teaching network educators answered the survey in October 24<sup>th</sup> and 27<sup>th</sup>. Apart from those teachers, nobody else answered (See Table I). Having encountered this inconvenience, the surveys were sent directly to graduated English teachers in the Commune of Valdivia via electronic mail. This time, most teachers answered back. Teachers who did not do it were visited at their schools.

It can be assumed that if teachers want to make changes in the educational system, they should be willing to participate in research studies. However, it can be inferred that this unwillingness is related to lack of time. Also, probably, teachers did not have internet at home or work, time to answer a 14 closed-ended survey, or interest in the assessment subject. This could be analyzed for further research.

As a final aspect, the results of this research concern the educational system, especially the assessment procedures in the teaching field. An overall evaluation of the Chilean educational system is needed, because educators need more hours to check tests to



English Teachers' Assessment Practices in Valdivia in the light of the Principle of Reliability.

avoid fatigue at the moment of correcting them. The State should increase and/or include hours for scoring tests because teachers use their own free time to do so. This not only will benefit teachers but also the educational system in general. There are many educators, schools and students who are aware of the overall benefits of assessment while several are not. The lack of knowledge and awareness about the process of assessing students by the educational community can make a great difference in the goals achieved by students from institutions in which assessment is considered as a source of feedback rather than a tool for measuring. As it was stated previously, the integral use of the assessment techniques mentioned reduces the gaps of students' achievement while recognizing individual and cultural differences as part of the teaching/learning process (Aslett, 2006; Butt, 2010; Saville-Troike, 1978). Additionally, educators are seen as facilitators in this process showing interest and caring about success in the classroom. Furthermore, assessment changes positively the nature of pupil-educator interaction (Brown, 2007). This change in the relationship is connected to the teacher, the student and the subject discipline helping and promoting effective practices in every educational institution.

## Conclusion

As Coombe *et al.* (2007) state, “reliability focuses on the empirical aspects of the measurement process” (p. xxiv). To measure this process –at least partially, – a survey was designed to evaluate reliability in the assessment practices of teachers in Valdivia. The research was focused on intra-rater reliability, specifically fatigue, negligence, and bias towards students. Results showed that, in general, this principle is met, except for one aspect that is not taken into consideration: Fatigue. The fact that EFL teachers do not check tests during daytime could affect intra-rater reliability, since problems due to fatigue and tiredness influence the scoring process (Brown, 2004; Coombe *et al.* 2007). In this regard, it is important to highlight that the reliability of test results influence the entire educational community, since teachers measure students' results and these result affect students' future. According to Robert Gordon, Thomas Kane and Douglas Staiger (2006), efficient teachers increment students' tests results and learning.

The second element involved in this study was test administration reliability, measured in respect to luminance, noise and seating arrangement (Brown, 2004). This principle was entirely met. The subjects commented that common sense was essential in order to facilitate students' performance and, therefore, results.

By applying the principles of assessment, educators and test developers can design fair tests which promote a positive view of the English language with which the students' efforts and strategies at the time of studying a foreign language, such as English, can be reflected in the scores obtained. As Areti Keramida & Iakovos Tsiplakides (2010) emphasize, “an important predictor of success in a foreign language is student's attitude towards it” (para.1). For this reason, the results presented in this research could help

English Teachers' Assessment Practices in Valdivia in the light of the Principle of Reliability.

educators consider tests as a tool to improve an underestimated aspect of the profession, such as the assessment process. In addition, EFL teachers can have an idea of the assessment practices in the Commune of Valdivia, and can thus draw generalizations about the current state of assessment in Chile.

To conclude, it is suggested that further research is carried out to investigate if English Foreign Language teachers have a schedule during working hours to score tests. This proposal should consider working hours, numbers of students per class, and amount of tests per semester. Additionally, the principle of assessment 'practicality': Time, resources, and testing process (Brown, 2004) may be studied, since this principle is intrinsically related to testing praxis. That is to say, it is an empirical procedure, much like reliability (Coombe *et al.*, 2007).

Finally, as it was suggested in the Discussion section, the lack of teachers' willingness to answer electronic surveys –and thus contribute with research studies in general– could be investigated, as it can be associated to their lack of time, fatigue after long days of work, but also to other factors which remain unknown. These and other topics, such as the real use of technology in the classroom, could be greatly beneficial to the current state of the quality of education in Chile.

## References

- Alexander, R. (2010). *The Cambridge Primary Review Research*. The University of Cambridge. Routledge: Taylor & Francis Group. Retrieved from [http://books.google.cl/books?id=y9wpQHUM1Q0C&pg=PA239&lpg=PA239&dq=shield+BM,+Dockrell+JE.&source=bl&ots=JQ4\\_O4hGtC&sig=LlSpXbsLeLgs6pMc6cQY\\_ydwmRU&hl=en&sa=X&ei=xbABUcCtLefB2QWqjYDQDQ&ved=0CFoQ6AEwBg#v=snippet&q=noise&f=false](http://books.google.cl/books?id=y9wpQHUM1Q0C&pg=PA239&lpg=PA239&dq=shield+BM,+Dockrell+JE.&source=bl&ots=JQ4_O4hGtC&sig=LlSpXbsLeLgs6pMc6cQY_ydwmRU&hl=en&sa=X&ei=xbABUcCtLefB2QWqjYDQDQ&ved=0CFoQ6AEwBg#v=snippet&q=noise&f=false)
- American Psychological Association. (2012). *Importance of Sleep*. Washington, DC, USA: American Psychological Association. Retrieved from <http://www.apa.org/topics/sleep/why.aspx>
- Arango, A. (2008). *The Failings of Chile's Education System: Institutionalized Inequality and a Preference for the Affluent*. Council on Hemispheric Affairs Research Associate: CoHa.Org. Retrieved from <http://www.coha.org/the-failings-of-chile%E2%80%99s-education-system-institutionalized-inequality-and-a-preference-for-the-affluent/>
- Aslett, H. (2006). *Reducing Variability, Increasing Reliability: Exploring the Psychology of Intra- and Inter-rater Reliability*. London Metropolitan University. UK: Department of Psychology, Investigations in University Teaching and Learning, 4 (1), ISSN 1740-5106. Retrieved from [http://www.londonmet.ac.uk/library/k36338\\_3.pdf](http://www.londonmet.ac.uk/library/k36338_3.pdf)
- Bachman, L. F., & Palmer, A. (1996). *Language Testing in Practice*. New York: Oxford University Press.
- Baldwin, S., McFadyen, L., & Stringer, E. (2010). *Integrating, Learning, and Action Research: Enhancing Instruction in the K-12 Classroom*. USA: Sage Publications, Inc.
- Bailey, K. (1994). *Methods of Social Research*. New York: The Free Press.

English Teachers' Assessment Practices in Valdivia in the light of the Principle of Reliability.

Barker, M., Farrelly, B., & Frederiks, F. (n.d). GIHE Good Practice Resource Booklet –

*Designing Culturally Inclusive Learning and Teaching Environments – Classroom Strategies*. Excerpt retrieved from [http://www.griffith.edu.au/\\_\\_data/assets/pdf\\_file/0011/184853/Creating-a-Culturally-Inclusive-Classroom-Environment-mcb2.pdf](http://www.griffith.edu.au/__data/assets/pdf_file/0011/184853/Creating-a-Culturally-Inclusive-Classroom-Environment-mcb2.pdf)

Borich, G., & Kubiszyn, T. (1999). *Educational Testing and Measurement: Classroom Application and Practices*. New Jersey: John Wiley & Sons, Inc.

Black, A., Lambert, N., & Rosenfield, P. (1985). *Desk Arrangement Effects on Pupil Classroom Behavior*. The American Psychological Association. *Journal of Educational Psychology* 1985, 77(1), 101-108. Retrieved from [http://www.summit.k12.co.us/cms/lib04/CO01001195/Centricity/Domain/72/Journal\\_of\\_Educational\\_Psychology.pdf](http://www.summit.k12.co.us/cms/lib04/CO01001195/Centricity/Domain/72/Journal_of_Educational_Psychology.pdf)

Blythe, T., & Associates. (1998). *The Teaching for Understanding Guide*. Active Learning Practices for Schools, *Teaching for Understanding: Ongoing Assessment*. Harvard University Graduate School of Education and Project Zero. Excerpt retrieved from <http://www.learner.org/workshops/socialstudies/pdf/session7/7.OngoingAssessment.pdf>

Brandsma, J. (n.d). *Education Equity and Social Exclusion: Final Synthesis Report*. Netherlands: University of Twente. Retrieved from [ftp://ftp.cordis.lu/pub/improving/docs/ser\\_clusters\\_education\\_social\\_exclusion\\_synthesis.pdf](ftp://ftp.cordis.lu/pub/improving/docs/ser_clusters_education_social_exclusion_synthesis.pdf)

Brown, D.H. (2004). *Language Assessment: Principles and Classroom Practices*. New York: Pearson Longman Press.

Brown, D.H. (2007). *Principles of Language Learning and Teaching*. (5<sup>th</sup> ed.). New York: Pearson Longman Press.

Butt, G. (2010). *Making Assessment Matter*. London: Continuum International Publishing Group.

English Teachers' Assessment Practices in Valdivia in the light of the Principle of Reliability.

Carnegie Mellon University. (n.d). *Learning Principles: Theory and Research- Based Principles of Learning*. Pittsburgh, USA: Carnegie Mellon University, Eberly Center for Teaching Excellence. Retrieved from <http://www.cmu.edu/teaching/principles/learning.html>

Coombe, C., Folse, K., & Hubley, N. (2007). *A practical guide to Assessing English Language Learners: The process of developing assessment*. Ann Arbor: The University of Michigan Press.

Crystal, D. (2003). *English as a Global Language*. Cambridge: Cambridge University Press.

Cruz, D. (2005). *Communication and Related Factors Affecting Academic Success among College Students*. The Florida State University: College of Communication. USA. Retrieved from <http://etd.lib.fsu.edu/theses/available/etd-11142005-153232/unrestricted/CruzMastersThesis.pdf>

Denton, P. (1992). *Seating Arrangement for Better Classroom Management*. Maryland, USA: General Conference of Seventh-day Adventists. The Journal of Adventist Education, 29-32. Retrieved from <http://circle.adventist.org/files/jae/en/jae199254052904.pdf>

Dockrell, J., & Shield, B. (2008 a). *The Effects of Classroom and Environmental Noise on Children's Academic Performance*. Retrieved from [http://www.icben.org/2008/PDFs/Shield\\_Dockrell.pdf](http://www.icben.org/2008/PDFs/Shield_Dockrell.pdf)

Dockrell, J., & Shield, B. (2008 b). *The Effects of Environmental and Classroom Noise on the Academic Attainments of Primary School Children*. United Kingdom: Faculty of

English Teachers' Assessment Practices in Valdivia in the light of the Principle of Reliability.

Engineering, Science and Built Environment, London South Bank University. Abstract  
retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/18177145>

Drave, N. (n.d). *Marker "fatigue" and Marking Reliability in Hong Kong's Language Proficiency Assessment for Teachers of English (LPATE)*. Retrieved from [http://www.iaea.info/documents/paper\\_30171b739.pdf](http://www.iaea.info/documents/paper_30171b739.pdf)

Educarchile. (2005, August 23). *El Inglés Abre Puertas*. Retrieved from <http://www.educarchile.cl/Portal.Base/Web/VerContenido.aspx?GUID=9d570195-30a3-4efa-9ac6-5b5982aad73c&ID=76723>

Educarchile. (2011, March 25). *Resultados SIMCE inglés 2010*. Retrieved from <http://www.educarchile.cl/Portal.Base/Web/VerContenido.aspx?ID=208390>

Education First. (n.d). *Education First English Proficiency Index [EF EPI]: Indice de Nivel de Ingles*. Retrieved from [http://www.ef-chile.cl/sitecore/\\_\\_\\_/~/media/efcom/epi/pdf/EF-EPI-2011-Spanish.pdf](http://www.ef-chile.cl/sitecore/___/~/media/efcom/epi/pdf/EF-EPI-2011-Spanish.pdf)

Educational Testing Service. (2003). *Linking Classroom Assessment with Student Learning*. Retrieved from [http://www.ets.org/Media/Tests/TOEFL\\_Institutional\\_Testing\\_Program/ELLM2002.pdf](http://www.ets.org/Media/Tests/TOEFL_Institutional_Testing_Program/ELLM2002.pdf)

Escamilla, K., & Grassi E. (2000). *Second Language Acquisition: A Brief Description of Second Language Acquisition*. USA: University of Colorado.

Ferrer, G. (2006). *Educational Assessment Systems in Latin America: Current Practice and Future Changes*. Retrieved from <http://www.uis.unesco.org/Education/Documents/Ferrer.pdf>

Figueras, N. (2005). *Testing, Testing, Everywhere, and not a While to Think*. *ELT Journal* 59 (1), 47-54.

English Teachers' Assessment Practices in Valdivia in the light of the Principle of Reliability.

Fowler, F. (1995). *Improving Survey Questions: Design and Evaluation*. (Vol. 38).

Thousand Oaks, CA: Sage Publications.

Gable, G. (1994). *Integrating case study and survey research methods: an example in information systems*. *European Journal of Information Systems*, 3(2), 112-126.

Gamoran, A. (2002). *Standards, Inequality and Ability Grouping in Schools*. Center for Educational Sociology, CES *Briefings*, N°25, September 2002. Retrieved from <http://www.ces.ed.ac.uk/PDF%20Files/Brief025.pdf>

Gedye, S. (n.d). *Formative Assessment and Feedback: A Review*. England, United Kingdom: Directorate of Teaching and Learning, University of Plymouth. Planet Issue, 23, 40-45. Retrieved from <http://www.gees.ac.uk/planet/p23/Gedye.pdf>

Gordon, R., Kane, T., & Staiger, D. (2006). *Identifying Effective Teachers Using Performance on the Job*. The Hamilton Project Advancing Opportunity, Prosperity and Growth. The Brookings Institution, discussion paper. Retrieved from [http://www.brookings.edu/views/Papers/200604hamilton\\_1.pdf](http://www.brookings.edu/views/Papers/200604hamilton_1.pdf)

Graesser, A. (2011). *Improving learning*. July 2011, 42 (7), 58, American Psychological Association. Retrieved from <http://www.apa.org/monitor/2011/07-08/ce-learning.aspx>

Greenstein, L. (2010). *What Teachers Really Need to Know About Formative Assessment*. USA, Alexandria, Virginia: Association for Supervision & Curriculum Development.

Griffith, G., & Mahadevan, S. (2006). *Sleep Deprivation Effect on Human Performance: A Meta-Analysis Approach*. Department of Energy National Laboratory, USA: Idaho National Laboratory (INL). Retrieved from <http://www.inl.gov/technicalpublications/documents/3394962.pdf>

Gunderson, L. (2008). *ESL (ELL) Literacy Instruction: A Guidebook to Theory and Practice*. (2<sup>nd</sup> ed.). New York: Routledge.



English Teachers' Assessment Practices in Valdivia in the light of the Principle of Reliability.

Gutierrez, C., & Peña, C. (2011). *The Use of Translation from a Communicative Perspective as an Effective Strategy to Teach and Learn Vocabulary in English*.

Retrieved from <http://cybertesis.uach.cl/tesis/uach/2011/ffg984u/doc/ffg984u.pdf>

Hanushek, E., Kain, J., & Rivkin, S. (2005). *Teachers, Schools, and Academic*

*Achievement*. *Econometrica*, 73(2), 417-458. Retrieved from

<http://www.economics.harvard.edu/faculty/staiger/files/HanushekRivkinKain%2BEcta%2B2005.pdf>

Harmer, J. (2007). *The Practice of English Language Teaching*. (4<sup>th</sup> ed.). Pearson Longman ELT.

Harris, M., & McCann, P. (1994). *Handbooks for the English Classrooms. Assessment*.

Oxford: Macmillan Heinemann English Language Teaching.

Haugen, L. (1999). *Classroom Assessment Techniques*. Center for Excellent in Learning and Teaching: Iowa State University, USA. Retrieved from <http://www.celt.iastate.edu/teaching/cat.html>

Hughes, A. (1989). *Testing for Language Teachers*. New York: Cambridge University Press.

Jackson, J., Smith, E., & Sparks, C. (2003). *Effects of Inequality and Reason for Inequality on Group Identification and Cooperation in Social Dilemmas*. *Group Processes & Intergroup Relations [GPIR]*, 6(2), 201-220. Retrieved from <http://www.indiana.edu/~smithlab/articles/ineqsocdil.pdf>

Keramida, A., & Tsiplakides, I. (2010). *Promoting Positive Attitude in ESL/EFL Classes*.

*The Internet TESL Journal*, 25 (1). Retrieved from [http://iteslj.org/Techniques/](http://iteslj.org/Techniques/Tsiplakides-PositiveAttitudes.html)

[Tsiplakides-PositiveAttitudes.html](http://iteslj.org/Techniques/Tsiplakides-PositiveAttitudes.html)

English Teachers' Assessment Practices in Valdivia in the light of the Principle of Reliability.

Kraemer, K. (1991). *The Information Systems Research Challenge: Survey Research Methods*. Vol. 3, Boston: Harvard Business School Press, USA.

Kraemer, K., & Pinsonneault, A. (1992). *Survey Research Methodology in Management Information Systems: An Assessment*. Retrieved from <http://staf.cs.ui.ac.id/WebKuliah/TKSI/Survey%20Method%20in%20MIS.pdf>

Krashen, S. (2009). *Principles and Practice in Second Language Acquisition*. Retrieved from [http://www.sdkrashen.com/Principles\\_and\\_Practice/index.html](http://www.sdkrashen.com/Principles_and_Practice/index.html) (Original work published 1982)

Lado, R. (1961). *Language Testing: The Construction and Use of Foreign Language Tests: A Teacher's Book*. London: Longman.

Lofty, N. (2012). *Seating Arrangement and Cooperative Learning Activities: Students' On-task/Off-task Participation in the EFL Classrooms*. The American University in Cairo, Egypt: School of Humanities and Social Sciences. Retrieved from <http://dar.aucegypt.edu/bitstream/handle/10526/3157/Classroom%20Seating%20arrangement%20and%20student%20participation.pdf>

Mandemach, B. J. (2003). *Faculty Resources Quick Tips: Formative Assessment in the Classroom*. Missouri: Park University, USA. Retrieved from <http://www.park.edu/cetl/quicktips/formative.html>

Manitoba Education, Citizenship and Youth. (2006). *Rethinking Classroom Assessment with Purpose in Mind: Assessing for Learning, Assessing as Learning, Assessment of Learning*. Winnipeg, Manitoba, Canada: MANITOBA Education [MECY]. Retrieved from [http://www.edu.gov.mb.ca/k12/assess/wncp/rethinking\\_assess\\_mb.pdf](http://www.edu.gov.mb.ca/k12/assess/wncp/rethinking_assess_mb.pdf)

English Teachers' Assessment Practices in Valdivia in the light of the Principle of Reliability.

McCreery, J., & Hill, T. (n.d). *Illuminating the Classroom Environment*. Retrieved from [http://www.lightingassociates.org/i/u/2127806/f/tech\\_sheets/Illuminating\\_the\\_Classroom\\_Environment.pdf](http://www.lightingassociates.org/i/u/2127806/f/tech_sheets/Illuminating_the_Classroom_Environment.pdf)

McNamara, T. (2004). Language Testing. In Davies, A. & Elder, C. *The Handbook of Applied Linguistics* (pp. 763-783). Oxford: Blackwell.

Nicol, D. (2007). *Principles of Good Assessment and Feedback: Theory and Practice*.

From the REAP International Online Conference on Assessment Design for Learners Responsibility, 29<sup>th</sup> – 31<sup>st</sup>, May 2007. Available at [http://www.york.ac.uk/media/staffhome/learningandteaching/documents/keyfactors/Principles\\_of\\_good\\_assessment\\_and\\_feedback.pdf](http://www.york.ac.uk/media/staffhome/learningandteaching/documents/keyfactors/Principles_of_good_assessment_and_feedback.pdf)

Northeast Energy Efficiency Partnership. (2002). *Energy Effective Lighting for Classrooms. Classroom Lighting KnowHow*. USA: DesignLights Consortium [NEEP]. Retrieved from [http://www.designlights.org/downloads/classroom\\_guide.pdf](http://www.designlights.org/downloads/classroom_guide.pdf)

Nova Scotia Department of Education (2001). *Bias Evaluation Instrument*. Province of Nova Scotia, Canada: Student Services. Retrieved from [http://studentservices.ednet.ns.ca/sites/default/files/bias\\_eval\\_ss.pdf](http://studentservices.ednet.ns.ca/sites/default/files/bias_eval_ss.pdf)

Nunan, D. (1992). *Program Evaluation: Research Methods in Language Learning*. New York: CUP, pp. 184-210.

Organization for Economic Co-Operation and Development. (2005). *Formative Assessment: Improving Learning in Secondary Classrooms*. OECD Publications. Retrieved from <http://www.oecd.org/edu/ceri/35661078.pdf>

Osram Sylvania Company. (2012). *Lighting Options by Space or Area: Classrooms*. Retrieved from <https://www.sylvania.com/en-us/applications/education/lighting-options-by-space-or-area/Pages/classrooms.aspx>

English Teachers' Assessment Practices in Valdivia in the light of the Principle of Reliability.

Palabra Maestra. (2011, January 06). Rodrigo Fábrega, Director del Programa Inglés Abre Puertas: *Los Maestros de Inglés: Los Alquimistas de la Educación en Chile*. Retrieved from <http://www.palabramaestra.premiocompartirmaestro.org/entrevistas.php?id=19>

Partin, R. (2009). *The Classroom Teachers' Survival Guide*. Practical Strategies, Management Techniques and Reproducibles for New and Experienced Teachers (3rd ed.). San Francisco, California, USA: Jossey- Bass.

Point Loma Nazarene University. (2012). *Classroom Culture that Inspires Student Learning*. San Diego, California, USA: Center for Teaching and Learning. Retrieved from <http://www.pointloma.edu/experience/academics/centers-institutes/center-teaching-learning/teaching-tips/classroom-culture-inspires-student-learning>

Popham, J. (2012). *Assessment Bias: How to Banish it. Mastering Assessment: A Self-Service System for Educators*. Massachusetts, USA: Pearson Education Inc. Retrieved from [http://ati.pearson.com/downloads/chapters/Popham\\_Bias\\_BK04.pdf](http://ati.pearson.com/downloads/chapters/Popham_Bias_BK04.pdf)

Philips Electronics Company. (2010). *Brighter Schools: Lighting a Sustainable Future for Education*. The Netherlands: Koninklijke Philips Electronics N.V. Retrieved from [http://www.lighting.philips.com/pwc\\_li/gb\\_en/connect/tools\\_literature/FINAL\\_Schools%20Segment%20Brochure\\_Int%20version%20june%202010.pdf](http://www.lighting.philips.com/pwc_li/gb_en/connect/tools_literature/FINAL_Schools%20Segment%20Brochure_Int%20version%20june%202010.pdf)

Ramsden, A. (1999). *Seating Arrangements*. USA: University of Delaware. Retrieved from <http://www4.uwsp.edu/education/pshaw/Seating%20Arrangements.htm>

Red Maestros de Maestros. (2004). *Resultados Nacionales del Diagnóstico en Inglés*. Retrieved from [http://www.rmm.cl/usuarios/jriva/doc/200609251337550.cambridge\\_resultados\\_simce.pdf](http://www.rmm.cl/usuarios/jriva/doc/200609251337550.cambridge_resultados_simce.pdf)

Richard-Amato, P. (2005). *Making it Happen: From Interactive to Participatory Language Teaching: Theory and Practice*. (3<sup>rd</sup> ed.). Boston: Longman Press.

English Teachers' Assessment Practices in Valdivia in the light of the Principle of Reliability.

Ruhl, K., & Wannarka, R. (2008). *Classroom Seating and Achievement: Seating Arrangements that Promote Positive Academic and Behavioral Outcomes: A Review of Empirical Research*. Oxford UK/ Massachusetts USA: Blackwell Publishing. Journal compilation NASEN, Support for learning, 23(2), 89-93. Retrieved from [http://www.academia.edu/706890/Seating\\_Arrangements\\_That\\_Promote\\_Positive\\_Academic\\_and\\_Behavioural\\_Outcomes\\_a\\_Review\\_of\\_Empirical\\_Research](http://www.academia.edu/706890/Seating_Arrangements_That_Promote_Positive_Academic_and_Behavioural_Outcomes_a_Review_of_Empirical_Research)

Saville-Troike, M. (1978). *A Guide to Culture in the Classroom*. National Clearinghouse for Bilingual Education.USA. Retrieved from <http://www.ncela.gwu.edu/files/rcd/BE000443/Culture.pdf>

Scantlebury, K. (2009). *Gender Bias in Teaching*. Education.com. The Gale Group, Inc. Retrieved from <http://www.education.com/reference/article/gender-bias-in-teaching/>

Stevens, R., Tyler, K., & Uqdah, A. (2009). *Cultural Bias in Teaching*. Education.com. The Gale Group, Inc. Retrieved from <http://www.education.com/reference/article/cultural-bias-in-teaching/>

Spiller, D. (2009). *Assessment: Feedback to Promote Students Learning*. New Zealand: The University of Waikato. Teaching Development/ Wāhanga Whakapakari Ako. Retrieved from [http://www.waikato.ac.nz/tdu/pdf/booklets/6\\_AssessmentFeedback.pdf](http://www.waikato.ac.nz/tdu/pdf/booklets/6_AssessmentFeedback.pdf)

The Texas University. (2011). *Assess Students: Why Assess Students Learning*. Austin, USA: The Texas University. Retrieved from <http://www.utexas.edu/academic/ctl/assessment/iar/students/plan/why.php>

UNESCO: International Institute for Educational Planning. (2005). *Educational Research: Some Basic Concepts and Terminology*. Paris, France: International Institute for Educational Planning/UNESCO. Retrieved from [http://www.iiep.unesco.org/fileadmin/user\\_upload/Cap\\_Dev\\_Training/Training\\_Materials/Quality/Qu\\_Mod1.pdf](http://www.iiep.unesco.org/fileadmin/user_upload/Cap_Dev_Training/Training_Materials/Quality/Qu_Mod1.pdf)

English Teachers' Assessment Practices in Valdivia in the light of the Principle of Reliability.

Ur, P. (1999). *A Course in Language Teaching: Practice and Theory: Trainee book*.

Cambridge: Cambridge University Press.

Weissglass, J. (2003). *Reason for Hope: You Can Change Educational Inequities*. The

National Coalition for Equity in Education [NCEE], University of California, Santa

Barbara. April 2003, 24-29. Retrieved from [http://ncee.education.ucsb.edu/articles/](http://ncee.education.ucsb.edu/articles/ReasonsForHope.pdf)

ReasonsForHope.pdf

Wilkins, A., & Winterbottom, M. (2009). *Lighting and Discomfort in the*

*Classroom*. Journal of Environmental Psychology, 29 (1), 63-75 DOI.

Woytek, A. (n.d). *Utilizing Assessment to Improve Student Motivation and Success*.

Chaminade University of Honolulu. Retrieved from [http://www.usca.edu/essays/](http://www.usca.edu/essays/vol142005/woytek.pdf)

vol142005/woytek.pdf

### **Appendix I: List of Educational Establishments**

1. Instituto Italia, Valdivia.
2. Liceo Polivalente Los Avellaños, Valdivia.
3. Instituto Comercial, Valdivia.
4. Instituto Superior de Administración y Turismo, Valdivia.
5. Liceo Santa María La Blanca, Valdivia.
6. Instituto Tecnológico del Sur, Valdivia.
7. Colegio Helvecia, Valdivia.
8. Liceo Técnico Profesional Mater Populi Dei, San José de la Mariquina.
9. Colegio José Manuel Balmaceda, Futrono.
10. Colegio Austral, Valdivia.
11. Colegio María Auxiliadora, Valdivia.
12. Colegio Alonso de Ercilla, Valdivia.
13. Instituto Salesiano, Valdivia.
14. Escuela España, Valdivia.
15. Escuela Proyecto de Futuro, Paillaco.
16. Colegio Laico, Valdivia.
17. Colegio San Lucas, Valdivia.
18. Colegio Particular Gracia y Paz, Valdivia.
19. Colegio Príncipe de Asturias, Valdivia.

## Appendix II: Survey

Estimada/o docente de Inglés,

Le invitamos cordialmente a participar en este **proyecto** liderado por **MS. Andrea Lizasoain**, el cual se llevará a cabo durante el segundo semestre 2012 en la comuna de Valdivia. Esta iniciativa busca recopilar información sobre el proceso de evaluación de la competencia lingüística en inglés de los escolares, el cual es de vital importancia dentro del ámbito pedagógico.

La presente **encuesta** consta de dos secciones, en la **sección I**, usted tiene que escribir los datos pedidos; y en la **sección II**, tiene que escoger una alternativa, en algunas preguntas se le pedirá escribir comentarios y/o especificaciones. Finalmente, cabe destacar que toda la información recopilada en esta encuesta es de **carácter anónimo** y es parte de **una tesis** para la obtención de grado.

De antemano muchas gracias por su apoyo,

### **Sección I: Identificación del docente**

1. Nombre:
2. Institución(es) en la que imparte clases:
3. Niveles en los que imparte clases:
4. Cantidad de alumnos por nivel o curso:

### **Sección II: Evaluaciones escritas**

1. Por lo general, ¿alcanzan los estudiantes a responder las pruebas escritas en el tiempo asignado? **Sí – No**
2. ¿Es incluido vocabulario nuevo en las instrucciones y/o preguntas de las pruebas?  
**Sí – No**
  - 2.1. Si respondió, "**Sí**", ¿explica usted el significado del vocabulario nuevo incluido en las pruebas antes de que los estudiantes comiencen a responder?  
**Sí – No**
3. Según su opinión, ¿realiza usted pruebas con contenidos enfocados a la realidad de los estudiantes? **Sí – No**
4. ¿Diseña usted pruebas usando ítems que siguen un orden lógico o secuencial?  
**Sí – No**
5. Según su horario, ¿en qué momento corrige usted las pruebas escritas con más frecuencia?
  - a) En la mañana
  - b) En la tarde
  - c) En la noche
6. Al momento de corregir pruebas escritas, ¿alguien lo/a ayuda? **Sí – No**
  - 6.1. Si su respuesta es "**Sí**", ¿quién lo/a ayuda?
7. ¿Considera usted el nombre de los estudiantes al revisar pruebas? **Sí – No**



**Comentarios:**

8. Al momento de corregir pruebas escritas, ¿hace usted algo para no ver el nombre de los estudiantes? **Sí – No**
9. Las instrucciones escritas de cada ítem de las pruebas son:
  - a) En Inglés
  - b) En Español
  - c) En ambos idiomas (Inglés y Español)
10. Las instrucciones orales de las pruebas escritas son:
  - a) En Español
  - b) En Inglés
  - c) En ambos idiomas (Español e Inglés)
11. ¿Diría usted que los estudiantes cuentan con suficiente luz para completar los ítems de las pruebas escritas? **Sí – No**
12. ¿Se preocupa usted de mantener una sala sin mayores ruidos externos? **Sí – No**
13. ¿Se preocupa usted de mantener un ambiente agradable al momento de tomar pruebas escritas? **Sí – No**
14. ¿Cómo organiza la sala de clases a la hora de tomar pruebas escritas?
  - a) Se hacen filas
  - b) Se deja un espacio entre estudiante y estudiante
  - c) No se realizan cambios
  - d) Otra (mencione):

**Muchas Gracias por contestar esta encuesta**

### **Appendix III: Survey Translation**

Esteemed English teacher,

We would like to cordially invite you to participate in this **Project**, headed by **MS. Andrea Lizasoain**, which will be carried out during the second semester of 2012 in the Commune of Valdivia. This initiative seeks to compile information regarding the process of evaluation of the linguistic competence in English of the students, which is of vital importance within the pedagogical sphere.

The attached survey consists of two sections. In **Section I**, you must fill in the required information; and in **Section II**, you must choose an option. In some questions, you will be asked to give comments and/or details. Finally, it must be stressed that all of the information collected in this survey will be treated in an anonymous manner and is to be part of a thesis necessary for the obtainment of a professional degree.

Thank you in advance for your support,

#### **Section I: Identification of teacher**

1. Name:
2. Institution(s) in which you give classes:
3. Levels in which you teach:
4. Amount of students per level or course:

#### **Section II: Written evaluations**

1. In general, do the students manage to complete written tests in the designated time?  
**Yes –No**
2. Is new vocabulary included in the tests' instructions and/or questions? **Yes –No**
  - 2.1. If you responded "**Yes**", do you explain the meaning of the new vocabulary included in the tests before the students start to answer?  
**Yes –No**
3. In your opinion, are the tests you set contextualized to the lives of the students?  
**Yes –No**
4. Do you design tests using items that follow a logical or sequential order?  
**Yes –No**
5. According to your schedule, at what time of the day do you usually correct written tests?
  - a) In the morning
  - b) In the evening
  - c) At night
6. While correcting written tests, does anyone help you? **Yes – No**
  - 6.1. If your answer is "**Yes**", who helps you?
7. Do you consider the names of the students when checking tests? **Yes –No**  
**Comments:**

8. While correcting written tests, is there anything you do to avoid seeing the names of the students? **Yes –No**
9. The written instructions for each item of the tests are:
  - a) In English
  - b) In Spanish
  - c) In both languages (English and Spanish)
10. The oral instructions for written tests are:
  - a) In Spanish
  - b) In English
  - c) In both languages (English and Spanish)
11. Would you say that the students have enough light to complete the items of the written tests? **Yes – No**
12. Do you make sure that the classroom is free from intrusive noise? **Yes –No**
13. Do you try to maintain a pleasant testing environment at the moment of applying written tests? **Yes –No**
14. How do you organize the classroom when you set written tests?
  - a) Make rows
  - b) Leave a space between students
  - c) No changes made
  - d) Other (mention):

**Thank you very much for answering this survey**