UNIVERSITY OF EL SALVADOR
SCHOOL OF ARTS AND SCIENCES
FOREIGN LANGUAGES DEPARTMENT

UNDERGRADUATE RESEARCH:

TESTING PRINCIPLES APPLICATION ON TESTS ADMINISTERED TO INTERMEDIATE INTENSIVE ENGLISH I STUDENTS OF THE FOREIGN LANGUAGE DEPARTMENT.

IN ORDER TO OBTAIN THE DEGREE OF:

LICENCIATURA EN IDIOMA INGLÉS OPCION ENSEÑANZA

PRESENTED BY:

NUBIA MARICELA AGUILAR GUEVARA   AG06013
ERICK ALEXANDER ARAGÓN CRUZ   AC06016
KARLA JEANNETTE FRANCO RIVERA   FR06007
CECILIA MIRANDA GIRÓN   MG06076

ADVISOR:

JUAN CARLOS CRUZ CUBÍAS, MA

SAN SALVADOR, EL SALVADOR, CENTRAL AMERICA, MAY 2012

AUTHORITIES OF THE UNIVERSITY OF EL SALVADOR

ING. MARIO ROBERTO NIETO LOVO
RECTOR

MTRA. ANA MARIA GLOWER DE ALVARADO
ACADEMIC VICE RECTOR

IN PROCESS TO BE ELECTED
ADMISTRATIVE VICE RECTOR

DRA. ANA LETICIA ZAVALETA DE AMAYA
SECRETARY GENERAL

AUTHORITIES OF SCHOOL OF ARTS AND SCIENCES

LIC. JOSE RAYMUNDO CALDERON MORAN
DEAN

MTRA. NORMA CECILIA BLANDON DE CASTRO
VICE-DEAN

MTRO. JULIO CESAR GRANDE RIVERA
SECRETARY

AUTHORITIES OF THE FOREIGN LANGUAGE DEPARTMENT

MTRO. JOSE RICARDO GAMERO ORTIZ
HEAD OF THE DEPARTMENT

RICARDO GARAY SALINAS, M Ed.
**GENERAL COORDINATOR OF
THE DEGREE PROCESSES**

JUAN CARLOS CRUZ CUBÍAS, M.A.
ADVISOR
CONTENTS

CONTENTS   **PAGE**

CHAPTER 4: ANALYSIS OF RESULTS

CHAPTER 5: CONCLUSIONS AND RECOMMENDATIONS

ANNEXES

INTRODUCTION

Assessment is one of the main components of learning and teaching process. It involves any formal or informal activity that is meant to collect information from learners to judge their language performance or language knowledge (Spratt, et al, 2005). In order to collect that information from students, teachers make use of some different tools like paper and pencil tests, role plays and tasks related to real life situations. The most common tool used to gather information is paper and pencil tests. Nevertheless, the information gathered from tests sometimes would not give valid information about students learning due to different factors that affect test design or development such as lack of time resources and the like.

Within the educational field tests serve as a tool used to make a decision (Pedhazur et al, 1991). This purpose rules every test, thus teachers are encouraged to determine what students' needs are and depending on those is how the test should be designed. Researchers discovered important elements that aid this process. Answering four important questions will help to design a quality test. For example, who will take the test (students), why is the test made (purpose), how will the test be evaluated (tasks and grades) and what will be tested (objectives or criteria) are all part of the development stages that a teacher should consider (Fleurquin, 2010). The quality of tests depends directly on its design. Consequently, the design of tests should follow some rules called testing principles in order to make it reliable, valid, useful and practical.

Therefore, recognizing what elements are involved in the test design process will make teachers highlight, search for specific information and concentrate on priorities. An important source that eases such process is the course syllabus. Not only because it offers the guidelines for the course but also for test content. In fact, tests must prove a bond with what has been shared in class and what is shown on tests (Kaplan & Saccuzzo, 2008). It is worthless to teach with the test when tests purpose is to measure what has been taught.

Along with the syllabus there are some guidelines that become important assets to design a test and that many teachers overlook. In the educational field reliability and validity are the most common testing principles; however, there are others of great importance as well. For instance, within validity: content, construct, criterion, face validity; student-related, teacher-related, administration reliability; authenticity, usefulness, practicality, washback. Applying these, "the cornerstones" (Combee & Hubly, 2007) guarantees better outcomes and an easy way of measuring what students have achieved.

Under no circumstance should a test be aimed to flunk students. On the contrary; teachers' task is to design the fairest test not only because it will benefit students but also because teachers will be able to draw conclusions from a valid test. And such will be only achieved if teachers pay close attention to what is placed on it, which in turn sets a proper expectation from students.

Having all this, the research project intended to determine to what extend were testing principles applied to the tests that were administered to student of the Intermediate English I of the semester II – 2011 from the Foreign Language Department. Also, it was intended to show and describe how these principles were applied to these tests. For this purpose, a focus group was made from a sample of Intermediate English I students. Besides, it was necessary to gather information from the teachers who taught this specific Intermediate English I courses in order to know further the design of the tests. To gather this information, a survey was passed along to teachers. All this information was valuable to offer recommendations for the application of testing principles and the improvement of testing experience for students.

CHAPTER 1

THE PROBLEM

1.1 STATEMENT OF THE PROBLEM

There is no doubt that tests are important tools to gather information about students' knowledge. Unfortunately, most of the times these instruments are not applied in an appropriate way by teachers since those tests do not fulfill some important principles of test design like reliability, validity, authenticity, practicality or usefulness. Making up good tests provides a guarantee of quality results not only for students but also for teachers. However, there is a certain lack of interest in revising each part of the tests to identify if it really complies with all the principles.

In the Foreign Language Department, the quality of tests administered to students has been questioned but never really investigated. Therefore, this research attempts to find out the extent of the use of testing principles on test administered to Intermediate English I students of the Foreign Language Department. The lack of research in this area makes this study more significant since it would be the antecedent for future projects.

1.2 FORMULATION OF THE PROBLEM

To what extent do teachers of the Foreign Language Department apply testing principles on tests administered to Intermediate English I students in the semester II - 2011?

1.3 OJECTIVES

General Objective

- To explore how the testing principles on tests administered to Intermediate English I students in the semester II - 2011 are applied.

Specific Objectives

1. To describe the application of testing principles on the exams in the Intermediate English I course in the Foreign Language Department.

2. To report how the application of testing principles improves the testing process.

3. To set the basis for future research in the testing process in the Foreign Language Department.

4. To offer recommendations for the application of testing principles in the exams administered in the Foreign Language Department.

1.4 RATIONALE FOR THE STUDY

Being testing the main tool used to gather information on what students know and can do; therefore, needs to be subject to close scrutiny on how it is designed and what is included. Teachers' judgments are based on students' responses yet to conclude if there is a progress or not, very important aspects to create a test must be considered. Such aspects have to do with testing principles guidance. For instance, considering reliability and validity is a good way of weighting what the course is aiming at, what is needed, what

has been taught among other important questions that could be easily answered by using testing principles at all times.

There is not much information about testing principles and testing in the Foreign Language Department. Indeed, research projects on these testing issues have not been carried out with students on the English Developmental Areas in the Department. Besides, since tests have become a major area of study within educational research, the study will try to find out more information on this topic focusing their efforts in the Foreign Language Department. In short, in the light of the lack of knowledge, this research aims to contribute to fill up this gap by gathering valuable information of application of testing principles in order to improve tests quality.

1.5 LIMITATIONS

One of the limitations encountered in this study was the dispositions to interview all the teachers from each group of the Intermediate English I. There were nine teachers who taught the ten groups of English. However, seven teachers were interviewed. One of these teachers was the advisor who could not have been interview due to the obvious reason of guiding the study. The other teacher could not be found in available time. Moreover, although there was ample literature about principles of assessment, the lack of interest on strictly analyzing each test and its parts was another constraint since there was a scarcity of studies done by other researchers. If there were more studies about the

application of testing principles, these findings could had been applied to the context of

the real application of these principles on the tests in the Foreign Language Department.

CHAPTER 2

THEORETICAL FRAMEWORK

2.1 BACKGROUND

Testing and Assessment

We live in a testing world. Worldwide testing is used to "label" people's knowledge. Tests are used for various educational and social purposes in society. For instance, people need to go on a test to get their driver's license. As well, in some countries people have to take a test in order to get their citizenship. Besides, most educational institutes require their applicants to take a general knowledge test. These examples as well as many others illustrate that testing plays an important role in society in everyday life.

Before starting, it is important to clarify some concepts. The concepts that are of main importance for this research are: assessment, evaluation and testing; also there is a need to explain how different they are. First of all, assessment involves any formal or informal activity that is meant to collect information from learners to judge their language performance or language knowledge (Spratt, Pulverness & Williams, 2005). As it has been described above, assessment is an extensive concept involving everything that teachers do to collect information (testing, random questions, oral presentations, tasks, etc.) to have insights about students' development and learning. In short, assessment is what teachers do to have an idea about students' learning and judge how well their learning is going.

Evaluation is the next concept that needs to be clarified, but what is the difference between assessment and evaluation? For this research evaluation will be defined as the

process of judging the students' learning base on an established standard or criteria and giving a value to that learning (Ontario Ministry of Education, 2010.) In other words, evaluation takes place after gathering data from students' learning and it judges what teachers gather through any assessment activity. In a few words, teachers in order to assess first gather data by testing students and after that they judge this data base on a certain criteria or standard as level of grammar use, spelling, fluency, accuracy and these judgments are what is called evaluation.

In the field of education, testing is especially important to provide valuable information whether the goals or objectives have been reached. Consequently, "Its most obvious function is to enable instructors to make judgments about the quality of student learning" (Piontek, 2008). In other words, testing is a tool that teachers use to gather valuable data about students' learning; therefore, if this is the case it could be said that testing is part of a bigger process that not only includes assigning grades and gathering information but also judging and analyzing students' learning progress (Kizlik, 2012) . Assessment is the big picture and both evaluation and testing are part of the process of assessing students. In summary, on one hand assessment is any activity conducted to know about students learning progress and it can be formal or informal and includes both gathering and judging data about student's learning. On the other hand, testing is the tool to collect data and evaluation the criteria to judge this data.

Due to the use of testing for gathering data in the last decades, tests have been seen by teachers and students as a way of getting scores, sometimes forgetting the big importance it has as a teaching tool and part of a bigger process. Testing has been one of

the most challenging and important stages in the learning process. "Testing is viewed as any of a variety of techniques that can capture what a person knows in response to a question" (Rudner & Schafer, 2002). With tests, it can be determined what the students currently know and can do. Independently of what tests are used for, there is no doubt that testing has been considered merely an evaluative instrument (McDaniel et, al 2007). The purpose of testing nowadays in most cases is to grade students or measure the quality of an institution; however, tests go beyond those objectives.

First and foremost, it must be essential knowing why assessing students is so important. Assessment comes in different sources to communicate to students what their weaknesses and areas of opportunities are, as well as their improvement throughout the course. These different sources might be feedback from their learning, tests, essays and the like (McAlpine, 2002).Taking tests as the method to gather information, to guarantee that this communication is effective and as honest as possible, it is important to take into account certain principles to make sure that tests are valid and reliable (U.S . Department of Labor, 1999). These principles must be held in mind by teachers whenever they construct a test, any type of test. The principles of testing fall under main aspects that make tests transparent and reliable to provide real expected targets.

Principles of Assessment

There is a variety of issues that must be taken into account when planning an assessment strategy and it is important to understand that assessment is a form of communication between all those involved in the teaching learning process (McAlpine,

2002). The importance of tests results is very significant; there is where some guidelines come into play to make tests more reliable and valid. Principles of assessment are rules that have to be followed by teachers when designing and administering a test. There is no doubt that these principles help teachers to enhance quality in the testing process. If they are ignored it cannot be said that a test is reliable, valid, authentic or useful. That is why this paper explains and clarifies the definition of each principle of assessment.

- **Validity and Reliability**

In spite of the fact that validity and reliability are two different things, McAlpine (2002) joins and define them as the overall quality of assessment. The relationship between them is that a valid test must be always reliable. That means that, teachers not only need to seek for reliability but also for validity on a test, in order to decide if a test is valid or reliable both must be present, so a test is not valid or vice versa . To define more deeply these principles we can say that validity refers to the capacity to measure what is supposed to measure and reliability in a test means giving the same results when a test is given to the same students more than once (McAlpine, 2002).

Types of Validity and Reliability

The University of Oregon (2005) has taken a close look into subdivisions regards to validity and reliability. Talking about test validity, we can say that content validity exists if what we are addressing and including in the test has been explained during instruction enhancing students familiarization with the information presented. Besides, it can be

stated that content validity is present by aligning the procedures used to teach such content during instruction as the guide used to design the tasks in a test (McAlpine, 2002).

1. Validity

In the field of education, validity is an important aspect that measures the learning outcomes through accurate conclusions based on the tests results. "Validity can be defined as the agreement between a test score or measure and the quality it is believed to measure." (Kaplan & Saccuzzo, 2008) When a test is valid on its basis, it has a clear purpose on what to measure and how it will do it. There are four main aspects of validity that determine the way the tests will be designed in order to provide accurate results depending on the skills that test measures. All the tests items need to be related with the purpose of the test. They also should reflect what the students expect from the tests. The following are the aspects that should be taken into account when creating a test.

1.1 Content Validity

"Content validity is a logical process where connections between the test items and the job-related tasks are established" (Professional Testing Inc. 2006) In simple words, content validity refers to the action of measuring what is supposed to be measured. When designing a test, teachers have to make sure that the items they are using in a test are closely related to the content seen in class. If this is not taken into account while designing a test, this test will fail to test what it is supposed to test because the topics are not related. This makes a test have poor or no content related evidence for validity

(Kaplan & Saccuzzo, 2008). In short, what teachers teach in classes must be what they include in the test otherwise there is not content validity.

1.2 Construct validity

Secondly, construct validity comes into play because it refers to the overall construct or trait being measured. For example if test –developers apply a communicative approach in class then the same theory or methodology should go hand by hand on the design of the test itself (Combee & Hubly, 2007) Similarly, if a test is overloaded, designed with one single item style or only assesses one aspect of the entire list of themes viewed in class this will also lack of construct validity. "Construct validity is essentially how closely the assessment relates to the domain you need to assess" (McAlpine, 2002). Whichever method and objectives used in class should be reflected throughout the test design; framing objectives will orient to stages of test development (Fleurquin, 2010) the who (students), why(purpose, objectives), what (method) and how (which test will be used).

1.3 Face Validity

"Face validity is concerned with whether a test looks as if it measures what it is supposed to measure but not really measure it at all" (Ley,2007). Therefore, face validity emerges from the student perspective thus pointing to what they expect to be included in a test, information like what type of contents, how many skills will there be assessed and what percentages each will receive (E-Teacher scholarship program, powerpoint, 2010). In short, face validity of a test is to look valid for the test takers, is the tests measures what students learnt but in a completely different way from the one it was taught. There is no

face validity because the test taker will feel that they are being tests different from what they were taught.

1.4 Criterion Validity

One more subarea of validity is criterion-related validity; it involves the degree of effectiveness with which performance on a test or procedure predicts performance. Therefore, criterion-related validity aims to demonstrate that grades are systematically related to criteria or outcome criterion (Cambridge ESOL Examinations, 2011). Criterion-related validity is not a new concept in the languages testing field, it can be tracked in time as far back as the 1950s (Lado, 1961). In short, criterion validity is about prediction rather than explanation, predicting something which is ought to be related, termed a criterion (Gabrenya Jr, 2003). Criterion validity evidence tells us just how well a test corresponds with a particular criterion (Kaplan & Saccuzzo, 2008). This means that according to the tests results, teachers can predict how well a student will perform in a certain task; if that student's grade is low then he or she will poorly perform on a task related to the content assessed, but if the grade is high, the performance will be quite good.

2. Reliability

In education tests are used to assign grades or marks to students' learning, but the accuracy of these grades or marks is not always satisfactory to student, parents or teachers. Tests always have certain discrepancies between real students' ability and the measurement of the ability (Kaplan & Saccuzzo, 2008). These discrepancies are seen as

errors of measurement, but this do not mean that tests are useless to measure students' ability. This is just inaccuracy that measurements have. Due to the existence of these errors of measurement, teachers have to look for ways to minimize these errors on their tests. This notion leads to the concept of reliability. What is reliability on tests? It can be defined as "the extent to which test are stable, consistent and free from errors of measurement" (Cambridge ESOL Examinations, 2011). Following this further, It can be said that "tests that are relatively free of measurement error are deemed to be reliable" (Kaplan & Saccuzzo, 2008). Reliability also can be affected for some factors, these factors are included as part of reliability achievement, as follows:

2.1 Administration Reliability

The administration reliability refers to the setting or the conditions under the tests are administered. According to Traub & Rowley (1991), some of the conditions that might reduce reliability on a test are physical conditions, instructions and time limits. Physical conditions refer specifically to the characteristics of the place where the test is being administered like the light, noise and levels of temperature of it. Besides, instructions given to students have to be considered. Teachers have to be careful when explaining or giving instructions to the examinees in order to avoid that they can guess the answers. The last but not less important condition is time limits. Teachers have to consider if the time is sufficient long for almost all students to finish the test. The lack of consideration of the factors described above could reduce considerably the reliability of the administration of the tests.

2.2 Student-related Reliability

Regarding the subdivisions of reliability, it is found student-related reliability; this factor takes into consideration emotional aspects like those that at very moment of being assessed were not present the last time the student took the same test (E-Teacher scholarship program, 2010). Among student- related reliability anxiety, temporary illness and fatigue can be also found. In fact, the influence feelings such anxiety, nervousness, headaches or simply a bad day can make scores differ from what the student is really able to do; therefore it reflects a delusive score and makes teachers draw different assumptions. Being assessed on several occasions; assessed when a person is prepared and best able to perform well, also ensures that students understand what is expected will avoid failing this principle (Coombe & Hubley; 2007).

2.3 Rater Reliability

Reliability is an important asset in education. It is specifically important in testing and it is shown in testing when results should produce the same scores on repeated trials. This is for example, several teachers grading an essay or an oral exam of the same student using the same rubric and therefore, providing similar scores. These raters are used when the type of test is based on subjectivity rather than objectivity but it requires a rating amount (Bresciani, Oakleaf, Kolkhorst, Nebeker, Barlow, Duncan & Hickmott, 2009). To avoid as much as possible subjectivity to provoke unfairness, rubrics are created to improve inter-rater reliability (when several teachers from an institution score the same

assignment). When these rubrics are established in the institutions and teachers are trained on how to score, the performance of students are more carefully assessed.

- **Authenticity**

"The history of language testing is, to a large extent, the history of attempts to bridge the gap between tests and real-life language use" (Ingram, 2003 pag. 3). These attempts, in other words, aim authenticity. This principle refers to how authentic the language of a test is; here teachers have to take into account if the language task they are using is related to real life situations. If test items are not contextualized and related to real situations they will be irrelevant and meaningless to students. Most tests contain multiple-choice, true and false, fill -in the blank and the like. Students are used to this kind of items and if a test does not have them it is an unusual test. However, "if teachers want to find out how well students perform on a task, they should put students to perform that task and observe how well he or she does it" (Doye, 1991). In brief, Authenticity in assessment requires students to do tasks that may replicate the use of language that they may encounter in real life situations and not tests isolated grammar patterns or language knowledge (Ingram, 2003).

- **Practicality**

In addition, practicality is also important since teachers need not only to be aware of what will be placed on the design but also to be aware of the means teacher will use on doing such test. Then a question arises and asks if our assessment meets our learners´ needs within available resources. By this is meant to find if we are counting economy of time, effort and money in testing so to simplify this answer teachers can break this down and check if there is no struggle on thinking how the design will be, no hassle on administering the test, no inconvenience grading and therefore on interpreting such results (Bachman & Palmer, 1996) hence, we are talking about three special aspects; human resources, material resources and time. Indeed, there should not be a lack of copies, tape recorders, seats, and desks; one that does not place an unreasonable demand on available resources (Cambridge ESOL examination, 2011).

- **Usefulness**

The usefulness of a test mainly depends on which purpose it is intended to fulfill as well of how aware students are of this reality and needs involved in this process. Everything has its specific purpose to be done. In testing, this is not the exception. Tests are useful as long as they provide information that leads to better predictions and understanding of students' learning progress (Kaplan & Saccuzzo, 2008). Every time teachers design a test, they should have in mind the targeted learners to be tested and the kind of test to be administered. The term usefulness applied to learning principles,

implies that both the teachers and students, and most importantly, the test, provides us with information about the class. Moreover, Bachman (2003) states that a fundamental matter of usefulness of tests in the classroom is that they help with decision making with evaluation and improvement of the educational program. Test usefulness requires that any language test must be developed with a specific purpose, a particular group of test takers and a specific language use in mind.

- **Washback**

One important fact is that tests also have an effect on students learning; they are powerful determiners of what happens in the classroom. This effect that can be negative or positive is called Washback. Depending on this effect, it can have a great impact on what students learn; teachers can decide what to teach and what improvements can be done on the teaching process (Phuong-Nga Nguyen, Griffin, Izard, 2008). When negative Washback is found on tests, teachers work thoroughly on their lesson and students work harder and take more seriously the subjects that are being tested.

2.2 THEORETICAL FOUNDATIONS

The Importance of Testing Principles

So far the history of testing principles has been explained, the criteria of gathering testing principles are vitally important in assessments. In fact, testing principles can be of

great help for teaching and for research. The importance held for teaching can be also found in the reason for designing assessments in our classrooms for these enhance motivation (Rudner, L. and Schafer, 2002) creates learning opportunities, gives feedback (students and teachers), provides grades, and serves as quality assurance mechanism (Oxfords Brooks University, 2002). Most of the time teachers are not aware how powerful to consider testing principles are and tend to overlook each one of our testing principles: reliability, practicality, validity, authenticity, usefulness, transparency.

Moreover, we can say that teachers can use testing principles for research purposes. All the reasons shared above are necessary yet the most important of all is that through using well –done assessments - that is by using testing principles – teachers are provided with feedback and so is also important to do research. For instance, the use of testing principles aids the research in the classrooms; it reveals valuable information on assessments strengths and also on the areas where it can be improved (Nicole & Macfarlane-Dick, 2004). In addition, testing principles may also reinstate assessment and teaching to the course objectives.

Testing Principles Application

Being aware of testing principles is just the beginning of the testing process, and what comes next is to apply these principles to have a good quality test. In order to apply all the testing principles that have been presented earlier, teachers must keep in mind that professional judgment is the foundation of assessment (McMillan & James H. 2000). This judgment is very important and it is used, for example: when constructing test question,

creating rubrics, scoring essays, and grading participation, to name a few. In other words, every activity related to assessment requires professional interpretation and decision making; moreover, assigning grades and assessing students' progress depends on the tests administrator appraisal and judgment. In short, if professional judgment is missing while making a test and interpreting its results, this test will fail to show accurate information about students' progress, and as a consequence, giving student feedback to improve their weakness becomes a harder task.

Assessment must be used more as a learning devise than a grading source. "Assessment is an integrated component of student learning experience" (assessment guide, 2008. page 11); this means that beside what is taught in classes students can learn even when they are being evaluated. The fact that assessment is mostly used for getting grades and judgments about students' progress is irrefutable, but this does not mean that assessment is not educative. In order to turn assessment into educative student must have a clear idea of what they are expected to learn and what they should be able to do in order to demonstrate what they have learned. In other words, assessment must have clear outcomes; clear learning outcome statements enable students to target their efforts appropriately (assessment guide, 2008). In brief, students can learn from assessment tasks if these (more than evaluate and give grades) are well designed and with clear purpose.

2.3 DEFINITIONS

- **Assessment** is a process by which information is obtained relative to some known objective or goal. Assessment is a broad term that includes testing (Kizlik, 2012). Assessment focuses on what students know, what they are able to do, and what values they have when they graduate. Assessment is concerned with the collective impact of a program on student learning.

- **Reliability** is the degree of consistency between two measures of the same thing. (Mehrens & Lehman,1987). Or it also can be defined as the measure of how stable, dependable, trustworthy, and consistent a test is in measuring the same thing each time (Worthen et al., 1993).

- **Validity** is the degree to which tests accomplish the purpose for which they are being used. (Worthen et al., 1993). Truthfulness: Does the test measure what it purports to measure? The extent to which certain inferences can be made from test scores or other measurement. (Mehrens and Lehman, 1987)

- **Authenticity** it involves real-life situations that are used to tests students' language abilities. Language learners are motivated to perform when they are faced with tasks that reflect real world situations and contexts (Coombe & Hubley,2007).

- **Construct** is the skill or proficiency an assessment is intended to measure.

- **Rubric** is the scoring criteria, scoring guide, rating scale and descriptors, or other framework used to evaluate responses.

- **Test** is a measurement device or technique used to quantify behavior or aid in the understanding and prediction of behavior.

- **Item** is a specific c stimulus to which a person responds overtly; this response can be scored or evaluated (for example, classified, graded on a scale, or counted). Because psychological and educational tests are made up of items, the data they produce are explicit and hence subject to scientific inquiry.

- **Focus group** research is a qualitative research method. Therefore, focus group is a social science research method (Morgan, D.L., 1993) it seeks to gather information that is beyond the scope of quantitative research email and telephone surveys and the like". The term "focus group" is often used to describe many types of group discussions (Grudens-Schuck, N. et al. 2012). Focus group research, however, is a true research method. As such, it uses a fairly standard methodology.

- **Principles** of testing mean the basic concepts and fundamental ideas that underlie all psychological and educational tests (Kaplan & Saccuzzo, 2008).

- A **convenience sample** is simply one where the *units* that are selected for inclusion in the *sample* are the *easiest to access*. A convenience sample is one of the main types of non-probability sampling methods. A convenience sample is made up of people who are easy to reach.

CHAPTER 3

METHODOLOGY

The purpose of this study was to explore the procedures held by teachers on the application of testing principles in the assessment practices in the Foreign Language Department. Having this intention, the research work aimed at obtaining specific information elicited from the Intermediate English setting. Thus, the work needed to provide textual descriptions of how students experienced this issue, where findings were applicable beyond the boundaries of the study; in other words, in qualitative research (Family Health International, 2000). Furthermore to interpret valuable findings, the need of qualitative methods such as structured conversations to obtain in-depth information or focus groups (Iowa State University, 2004) and interviews took place. Students and teachers in the Intermediate English I courses were involved in the research. These two sources helped to have a better understanding and an overall picture of what students experience and what teachers do regarding testing.

3.1 RESEARCH LEVEL

The research framed the testing environment from its top down to its end, the application. Searching information from how teachers design the tests and what students encounter when being evaluated. Due to these aspects the research work started the study from the testing principles view. Since this area has been superficially criticized by

teachers and students yet no study has been done to find out what exactly happens; this research sought to reveal findings and suggestions for testing in Intermediate English I courses in the Foreign Language Department.

3.2 RESEARCH DESIGN

There is a need to determine the application of testing principles on the test administered to Intermediate English students; consequently, in order to have the population insight the study was developed under a qualitative research basis. Using a qualitative method enhances the research purpose for collecting evidence on values, opinions and behaviors among students, teachers and tests themselves. In order to obtain specific type of data for a qualitative method, special treatment and means were applied: focus group and in-depth interviews. In this way, focus group was used to gather students' opinions and in-depth interviews were used for teachers.

3.3 PARTICIPANTS/SAMPLE

The study focused on two types of participants, students and teachers from the Intermediate Intensive English I courses of the Foreign Language Department at the University of El Salvador. This study aimed to reveal to what extent testing principles are applied on test administered to Intermediate English I students of the Foreign Language Department. Teachers and students from the ten Intermediate English courses were

selected since at this level both teachers and students establish the bases of the whole teaching and learning language process. Convenience sampling; sampling chose at the convenience of the researchers (Fairfax County Department of Systems Management for Human Services, 2003) was used to select students and teachers as the main participants in this study because of two main reasons; first, researchers had easy access to them and second because researchers had the opportunity to choose the high achievers from each course in order to get significant information for the focus group. Therefore, researchers had access to seven of nine teachers of the Intermediate English courses who had the best of intentions to be interviewed.

3.4 DATA COLLECTION

a) Intervention/Materials

First and foremost, the focus group technique was used to generate valid information from a group interview which had selected different students from each of the English intermediate I courses. Since focus group is a session that inquires the human perspective, a way of revealing data on beliefs, opinions and values (Family Health International, 2000); worksheets were given with questions related to testing within the English courses. Questions gathered on each worksheet had been based on one of the testing principles and were open ended giving the freedom to analyze, compare and share ideas. Also, pupils were asked to write answers down and share them with the rest of the group to start a discussion, yet some follow up questions were created depending on their

responses to keep the discussion going. Likewise, to save this valuable information the use of technology such as laptop computers, projector, video camera and photo camera was needed. Rather than only limiting responses written on sheets; in order to make it more valid and to document the research work, the focus group was videotaped recorded and teachers' interviews were audio recorded at the same time the ones written down were answered in computers and saved for references.

b) Measurement/Instrument.

So as to evaluate all these pinions and behaviors the research work required some means such as comparison tables, rubrics, worksheets, exams and the course syllabus. Since the study wanted to supplement the knowledge gained in the focus group with the ones given on interviews; research organized questions that both parties had in common and its answers were put next to each other in a comparison table. Moreover, the study needed to have a rubric so that testing principles could be checked within mid-terms and final exams; and then grade them in relation with a rubric of the study design that rated exams with: high, mid, low testing principles application fulfillment. Similarly, the course syllabus worked hand by hand with exams rubric; in fact, it served as a guideline to compare what teacher responses were on test design along with the syllabus and finally comparing those findings with what exams really gathered. Finally, questions asked on worksheets were paired depending on one testing principle. Pupils could discuss, analyze and share point of views since these were open-ended questions.

3.5   DATA ANALYSIS

a) Data Collection

In order to obtain the information necessary for this study, data were collected in the following manner: Firstly, it was decided to pass along an interview to teachers of the Intermediate English I courses, which was revised and validated by the advisor. To have the validation of the question from the interview, they were submitted to an analysis by the advisor in order to verify that the questions that would be addressed to teachers were conveying the correct aspects to gather information about the principles of assessment used in the tests. Secondly, since there was also needed to have valuable information from students as well, it was decided to have a focus group with a convenience sampling of students, who were carefully selected from the ten groups of Intermediate English from the semester II-2011. These students selected were the top five from each group, in order to have significant information from them. The selection of the students was made with the help of the teacher in charge of each group.

b) Data Analysis Procedure

Since this study gathered qualitative data from teachers and students, a triangulation analysis was used in order to linkage the results from the three main components of the research: students, teachers and theory. Information from teachers was compared with the information from the students and connected with the theoretical

framework. Furthermore, the analysis obtained from the triangulation was associated with the analysis of the exams in order to obtain substantial evidence that the exams do contain or not the principles of assessment versus the information collected from teachers and students.   Analyzing and comparing the information from three sources gave this research a whole view of testing principles application.

CHAPTER 4

RESULTS AND ANALYSIS OF RESULTS

4.1 Principles of Assessment application on tests in the Foreign language Department

Triangulation Analysis:

Teachers interviews, students' points of view and theory

Revising opinions and answers from teachers and students is crucial to a study; such is the case of this project, thus great relevance is given to how those answers were made. As previously mentioned, principles of assessment are rules that have to be followed by teachers when designing and administering a test (McAlpine & Clark, 2002); therefore, the following information reveals not only evidence provided by teachers but also that from students which will be compared on the bases of principles of assessment in order to identify the extent to which these principles are applied in the Intermediate English I courses in the Foreign Language Department. After organizing and analyzing the data collected from the interviews to teachers and the focus group to the students, it could be seen that there is a significant difference in the point of view on how the principles are applied to the tests.

1.   **Principle of Validity**

The principle of validity is sub-divided into five aspects which tell how well the test is structured to measure what it is supposed to measure. It was found important evidence about these five main types of validity from teachers and students.

Construct Validity

It refers to the overall construct or trait being measured (Combee &Hubly, 2007). The first question asked to teachers and students said: Do you consider that test is well designed to measure class objectives? Here, most teachers' answers were that they took into account class objectives to design their tests. On the other hand, students' answers were that they are not aware of the class objectives and what they were expected to do at the end of the course. They do not really know if tests measure class objectives. Students might have the syllabus, but this was not thoroughly explained so that they could have an overall picture from where they were coming and what level or abilities. Then they will be able to reach and have a broad view on how well tests are designed to measure these objectives.

Another question that arouse for construct validity was: Do the tests administered in Intermediate Intensive English I evaluate all the linguistic competences students are supposed to achieve? Teachers answered that they are focusing on oral performance while students said that there are some skills that receive more importance than others in our course. Thus at the end it is difficult to state if we are gathering all the goals of the course. Teachers are oriented on what skills they want to develop, improve and take care

of yet students are worried because they focus on the skills that have more percentage still do not know if at the end of the course having a high score as demanded is proof that all the linguistic competences have been achieved. A result given by not making students aware at the beginning of what linguistic competences an intermediate level I will be achieving.

Face Validity

Face validity takes place when a test looks as if it measures what it is supposed to measure (Ley, 2007). Therefore, one of the aspects that face validity involves is the information that teachers give to the students before taking or administering the test. Taking this aspect into account, researchers asked both teachers and students the following question: Before carrying out a test, are students told in advance what the test is going to contain? Teachers and students agreed that exams contain face validity, since students are told in advance what is going to be evaluated in the test and when it is going to take place.

Content Validity

A test, in order to have content validity, should include what has been explained during instruction. Teachers have to make sure that the items they are using in a test are closely related to the content seen in class (Professional Testing Inc. 2006). On this area, teachers and students were asked to explain about the contents included in the tests. After comparing and analyzing the answer from teachers and students to the question: How closely related are test items to the topics taught in classes? Teachers agreed that

what was on the tests was what they have taught and the class and the questions were similar to what they have practiced with students. Students claimed the majority of items are related to what they do in class. However, in a few of questions there was no relation between what they practice in class and what they are asked to do in some items.

In addition, there was another important question added to the content validity principle asked to students and teachers as well: Are topics that have not been taught in class included in the test? Teachers said that it was unfair to include topics not studied at all in class; nevertheless, teachers pinpointed that in some cases they included topics that were not taught in class but they let students know in advance so that students could study and get prepared for the exam. However, Students advocated that sometimes they were not told exactly what topics would be included and that studying by their own is much difficult than discussing topics in class and leaves gaps. When teachers assign topics for students in order to be self-studied, teachers should give feedback and make sure the topics are understood before making the exam. In this way teachers will fulfill content validity and avoid students having any type of surprises.

Criterion Validity

Criterion validity tells us just how well a test corresponds with a particular criterion (Kaplan & Saccuzzo, 2008). This means that according to the tests results teachers can predict how well a student will perform in a certain task; if that student's grade is low then he or she will poorly perform on a task related to the content assessed, but if the grade is high the performance will be quite good. Therefore, in order to investigate further

whether tests administered in the Intermediate English I level contain criterion validity, it was asked to students and teachers the types of items that are best for showing language abilities. They were given to choose between supply (open-ended questions, fill in the blanks) and selection (multiple choice, matching) items.

Teachers stated that supply items are the best that show language abilities since they make students analyze whereas with selection items it is easier to cheat and there is more chance to guess the answer. For students on the other hand, selection items are the best since they have the opportunity to analyze the answer rather than memorizing pieces or details of information.

2.   Principle of Reliability

The principle of reliability defines the extent to which tests are consistent and free from error of measurement. It has to be looked for ways to minimize these errors on tests. On this research, was found important information about administration reliability on the tests in the Intermediate English I in the Foreign Language Department those findings are explained through the two triangulations. The first triangulation took into account administration due to the types of questions answered just by teachers and students, the other features of reliability are explained next in the second triangulation.

Administration Reliability

A test is considered reliable if it has consistency and if it is free from errors of measurement. Nevertheless, according to Traub & Rowley (1991) the reliability of a test would be reduced because of certain conditions. One of them is physical conditions which refer to the place where the test is administered. Students and teachers of Foreign Language Department of the University of El Salvador were asked if they consider the settings and the assets from the department are adequate to administer the test. After collecting their answers, researchers found out that both teachers and students consider that the conditions are not appropriate for taking and administering a test since there are not only much interference around the Foreign Language Department but also the space and the amounts of desks are not enough. At this point, it can be said that the reliability of the tests taken and administered in this department is noticeably reduced.

3. **Principle of Authenticity**

In order for tests to comply with the authenticity principle, they have to be related to real life situations. This means that tests must be contextualized according to the students' atmosphere and situations to be relevant and meaningful to them. Students should be required to do tasks that replicate the uses of language in real life (Ingram, 2003). In order to provide information whether this principles is applied on the tests, the following question was asked to teachers and students as well: In what ways are realistic uses of the language included on the tests? In this case, teachers state that they try to use

realistic uses of the language with task-based approaches, with situations concerning their real context. On the other hand, students agreed that most of the times, tests evaluate but isolated pieces of vocabulary and structures, which are not related at all with their context. Some discrepancy is found on between teachers and students; however, both parts agree that tests should be improved on the contextualization of language in real life.

4.   Principle of Usefulness

To be aware of the intended purpose of the test, it is fundamental to achieve the goals established for the course. This principle of usefulness mainly depends on the purpose of the test. Likewise, the fundamental ingredient of usefulness is that it helps us with the improvement of decision making and evaluation in the class (Bachman, 2003). In order to investigate further if the tests administered to Intermediate Intensive English students comply with the principle of usefulness, students and teachers were asked the following question: How useful are test results to measure students' learning? Teachers agreed that tests are very useful since they reflect what students have achieved. It reflects what teachers are doing in class and how results helps them to make decisions about what should be done in class and to achieve the targeted purpose.

Similarly, students stated that test results provide important information about how well they are doing in their learning process and throughout the course. It seems therefore that both parts are aware of the usefulness that tests are given in the Foreign language Department.

5.   Washback

Beyond measuring what students have learned and achieved during the course, tests have also important effects on students' learning. These effects can be either positive or negative and are powerful determiners of what happens in the classroom (Nguyen, Griffin, Izard, 2008). Depending on this effect, teachers can decide what can be done to improve teaching. Teachers from the Intermediate English I said that when analyzing the tests results, they see the mistakes which they reflect on and become aware that the topic was not taught well or it was not clear for students; therefore, they can make a review or do something to improve. Regarding students' judgments, the results obtained on the tests are closely related with the importance they give to the course itself. For students, it was asked in what ways the grades and feedback provided affected their learning. They said that if the grades obtained are acceptable, they will be motivated to improve their performance. In a nutshell, it can be seen from this perspective, that students and teachers feel motivated to improve their performance and teaching as well, when the results of the tests are the ones as desired.

4.2 Principles of Assessment application on tests in the Foreign language Department

Triangulation Analysis:

Test design, teachers' responses and theory

Besides the valuable opinions and points of view from teachers and students about testing and test design in the Foreign Language Department, it must be also considered analyzing the tests themselves and comparing them to how they should be designed according to the level of the English course. In fact, findings should reveal a link between teachers' answers and the test organization. Therefore, in order to find out these insights, a rubric of the researchers design based on the theory of testing principles was created to analyze and verify if those exams fulfill the requirements of a high quality exam (the rubric was created by the researchers based on authors like McAlpine (2002), Fleurquin (2010) , Kaplan & Saccuzzo (2008) among others). Then, following analysis is based on the theory of testing principles, teacher's responses, the syllabus of the course and the researchers' findings. For further reference, teacher's responses, syllabus and rubrics can be found in the annexes.

1. **Listening Exam**

   (Mid-term and Final)

The following analysis was made taking into account the principles of assessment and whether they apply to the listening mid-term and final exams from the Intermediate English I courses administered in the Foreign Language Department. Valuable findings from each principle could be seen with a slight difference from teachers' responses and the tests themselves.

Principle of Validity

- **Construct Validity**

Construct validity has to be with course objectives and methodology. According to Fleurquin (2010), objectives of the teaching method should be reflected throughout the test design. Based on this, researchers asked teachers of the Foreign Language Department if their tests are designed to measure class objectives, the majority of teachers said that objectives are always taken into account when designing their tests. Nevertheless, researchers found out that the listening exams are partially fulfilling class objectives because some are in a way impossible to reach just by doing a multiple choice and fill in the blanks tests. Therefore, the objectives established in the syllabus for English Intermediate I were not completely reflected on the listening tests subsequently, those tests lack of construct validity.

- **Content Validity**

Robert M. Kaplan and Dennis P. Saccuzzo (2008) agree that a test which has poor or any content related to the class will be not valid. Researcher used this foundation to analyze if listening exams administered to Intermediate students are closely related with the topics taught in class. The exams were compared with the syllabus which describes the topics seen in class and found out that some topics testing in the listening part are not related with the topics of the class in class. Nevertheless, teachers' responses to the question: Are topics that have not been taught in class included in the test? Is the opposite of the researcher's findings saying that topics are related.

- **Face Validity**

Ley (2007) defines face validity as the quality of a test that it measures what is supposed to measure. To analyze this point, researchers took closely attention to test design, even though face validity emerges from student's perspectives researchers took into account teacher's responses about their testing design process and also the test itself, in order to prove if it fulfills some features of the face validity principle. The first aspect taken into account to verify if the test has face validity was the format used by the teachers of intermediate English, comparisons between the mid-term and the final listening test were made and researchers found out that the structure of them were similar both had multiple choice and fill in the blanks items. Besides, researchers sought for the organization of the test and it was proved that both tests were logically organized. Finally the level of difficulty of the test was taken into account and it seemed to be a little

bit challenging for students' English level. Nonetheless, the listening tests fulfill the basic features of face validity.

Principle of Authenticity

The principle of authenticity seeks to establish real life situations on tests. According to Wiggins (1990) authentic assessment is designed to demonstrate whether the student can generate full and valid answers in relation to the task or challenge at hand. In order to analyze if teachers of the Foreign Language Department are using authentic items on the listening tests, researchers checked first, if the test included real-life scenarios and it was discovered that both tests had situations that students could face in real life like a job interview, hotel check-in and college life. This aspect made the items of the exam meaningful for students. Another aspect was if the test had natural language, it was seen that teachers used recordings of native English speakers for the listening to provide that natural language. Finally, teachers were asked if realistic language is evaluated in their tests their responses were that they tried to use it as much as possible. In conclusion, the listening part of the mid-term and final Intermediate English tests accomplished the main features of authenticity.

Principle of Practicality

Practicality in a test refers to a test that is not that charge with questions, if it too long or if it is too short, if it is easy to grade and appropriate for the student (Bachman & Palmer, 1996). The aspects used to verify if the listening tests were practical were, first, if the instructions were clear. Both test mid-term and final had unambiguous instructions that were simple, short and understandable for students. A second criterion was the relationship between the amount of items and the available time for the tests. The mid-term test had lesser items than the final test; however, the audio were longer and more challenging for students while the final test had more items to answer and more short conversations. The time was well distributed to complete the exam; nonetheless, there was not enough time for students to read the options before listening and answering the test. The last criterion was if the test had a scoring procedure that is specific and time-efficient. According to teachers responses, scoring is a time consuming activity, however the types of items that the listening tests had are not time consuming, and it is a lot of easier for teachers to score them. Taking those features in consideration, the listening part of the tests of Intermediate English I had the sense of practicality not only for teachers but also students.

2. **Grammar Exam**

(Mid-term and Final)

In order to find out whether the tests do comply with the principles of assessment regardless students and teachers opinions, a triangulation analysis was made to compare the tests themselves with the syllabus and the teachers' interviews. This analysis helped the researchers to find valuable information about the structure and elaboration of tests.

Principle of Validity

- **Construct Validity**

In the Intermediate English I syllabus course, the methodology described to be used was a communicative methodology. A test in order to have construct validity, the design of it must be connected with the methodology used in class and should be fully linked with the themes and class activities (Combee & Hubly, 2007). Based on this principle, it was found that in the Grammar tests the items were highly attached to the classroom objectives. This goes accordingly to what teachers said about trying to make as much as possible to fulfill the objectives within the tests. Furthermore, the design of the tests was congruent with the methodology since it went accordingly to the class activities. Nevertheless, these tests did not allow students to perform with the language since all the items are mainly made as structures, where students do not really implement the language. This implicates that it is not really known whether students can use language on

real life situations. They might know how to use structures, but not how to use it in real life context.

- **Content Validity**

The course syllabus is a guide where the objectives and topics to be seen in class are achieved. During the course some topics are studied while others are left aside for any reason. When it comes the time for testing, the test items should contain and should be closely related to what the content seen in class was (Kaplan & Saccuzzo, 2008). If the topics on the tests are not related with the topics seen in class, it fails to tests what it's supposed to test. In the Intermediate English I tests was found that some topics were added in the mid-term before they were studied. However, these same topics were included in the final exam. This is a little contrary to what teachers said about not adding topics not seen in class. Another important aspect about content validity is that these tests do not really assess students' ability to speak a second language since the settings of the items do not apply at all to the students' conversational context.

- **Face Validity**

When students see a test, they expect to see the same structure pattern that they have already been given and the contents already studied (Ley, 2007). Therefore, face validity in a test is what it is expected to be evaluated. The Intermediate English I exams use the appropriate level of difficulty from vocabulary and grammar structures. Besides, the format of the test is similar to rest of the tests. For that reason, students are aware of a pattern to which they are familiar with. However, the structure of the tests is not

logically organized at all since it does not follow a sequence where easy structures come before complex grammar structures.

Principle of Authenticity

Grammar tests in the Intermediate English I courses should attempt to relate their contents to real life situations. If they authentic, they should have language contextualized and related to real life scenarios that are relevant and meaningful to students (Ingram, 2003). The test items used in these exams are not associated to real situations that students could have in their context. Even though the grammar structures are well applied, the items somehow do not show meaningful use of it in the context where they have been selected. It is a context where students do not really match with theirs.

Principle of Practicality

In order for exams to measure what they are supposed to measure they should be practical in time, effort and money as to simplify the task for design and grading (Bachman & Palmer, 1996). It was found that tests from the Intermediate English I are quite aware of the practicality principle. The instructions of the tests are clearly stated and therefore there were not found instructions that were ambiguous. Additionally, the amount of items was proportional to the time frame that students were given. Finally, the scoring procedure was specific and time efficient since all the items were specifically multiple

choice and fill in the blanks. For teachers these are the best types of items since they do not involve a complex scoring procedure.

3. Oral Exam

(Mid-term and Final)

The following analysis embraces four important sources for oral exams: mid-term and final exam, syllabus, teachers' responses on test design and testing principles theory. The analyses follow this way the first two; mid-term and final oral exam were rated with the rubric based on the testing principles and syllabus, finally a comparison among teachers responses.

Principle of Validity

- **Construct validity**

One of the construct validity criteria stated: The test design is congruent to the class communicative methodology, and this insight was given. Both mid-term and final oral exam achieved this for the test design enables to talk by giving the broad scenario. Then students were in charge of advocating or contradicting a topic; same methodology used in class when having discussions, so it can be said that teachers respected what was stated in the syllabus. As they explained, objectives and methodology have a connection (Combee & Hubly, 2007)   Along with this criterion the rubric had: classroom objectives are

identified and appropriately framed. The two exams were very similar and pursued course objectives which intention were to make students share, comment and discuss; however, since the final oral exam provided a list of questions not all fulfilled this aim for there were close-ended questions. Teachers said they chose items where students had the freedom to create and perform with the language yet not all of the questions gave freedom because their answers were to list or name only.

Moreover, when rating, test allows students to perform with the language. Even though in the mid-term oral exam test allows students to prepare themselves to pose a stand and to convey ideas, there were some questions on the final which answers purposes are of listing names, thus the item will not develop a communicative competence fully. This differs from what teachers seek but it is important to mention that most of the questions covered this purpose

- **Content Validity**

For content validity the following was stated: The topics taught in class are reflected on the test, here there was a variation showing. For instance, in the final oral exam topics are closely related to the ones studied in class and questions are similar to those held in discussions. While in the mid-term exam not all topics were taught in class and are not in the syllabus, some were chosen because they enabled discussions. This could in a way fail the connection between what was seen in class and what was in the exam (Kaplan & Saccuzzo, 2008).   Similarly, this statement was discussed: The test assesses student´s

ability to speak a second language in a conversational setting. Here, there was a slight difference found. By conversational setting it is assume that there will be the interference of two or more individuals, in the mid-term exam students prepared themselves on a topic and pose their stand yet teachers did not interfere in this process. While in the final exam, students were able to arrange and link ideas regarding a topic for teachers were involved, they asked follow up questions depending on students' answers; this made a more conversational flow.

- **Face validity**

Also for face validity the following was mentioned: Tests format is similar to the rest of the test again very important aspects emerged from both exams. For the mid-term exam, students only had a list of topics where no questions were listed whereas in the final teachers ease the exam by giving a list of questions, students could easily prepare themselves to answer and outline ideas or opinions (E-Teacher scholarship program, powerpoint, 2010). Moving on with face validity, this point was also examined: Test difficulty is appropriately to the students' English level. Both the mid-term exam and final exam were appropriate since students knew the topics also had the opportunity of developing at their own pace and level.

Teachers are then considering their audience level (E-Teacher scholarship program, powerpoint, 2010).   In addition to this principle, when organization came into

play some important aspects were found with: The structure of the test is organized logically. Firstly, in the mid-term exam Topics did not follow an order from how they were taught, however; all the topics were grouped depending on the problem. While in the final, each set of questions were grouped depending on its units in the syllabus and highlights the number of the unit from where questions were made. Teachers explained that diversity was due to how skills are split therefore different designs are gathered. This point is very important because is the degree to which a test looks right for a student.

Principle of Authenticity

Additionally, there were points involving authenticity such as: Tests are related to real-life scenarios. As a result, in the mid-term exam most of the topics came from real-life scenarios (Ingram, 2003 pag. 3); however, there were a few topics which are not well known for students. When checking the final exam, all three topics: best days of your life, screen test, crime doesn't pay; gathered in the test were related to real-life scenarios. As previously mentioned there were some topics included because it seem appealing for discussion purposes yet these topics were not listed in the syllabus and were not closely related to our environment. This is teachers concern to keep exams contextualized.

Besides, the study needed to find out how authentic tests were so the following was studied: Test items are authentic and meaningful. By reviewing the mid-term exam, it was found a meaningful connection between those topics for students are used to hear, see or live with yet having to talk about an odd topic will not make it authentic. The

contrary happened in the final where students had topics things they can encounter in everyday life they would easily make connections. Teacher said that once they finish the test, all teachers review the final work and made adjustments were possible in order for it to be authentic. Another aspect for authenticity was: Language used on test is natural. Vocabulary and tenses used to ask students suited their level so that they could understand what was being asked on both exams.

Principle of Practicality

To get information applied for practicality the following statement was made: The test instructions are clearly stated. By reviewing both exams it was confirmed how well and clear instructions were; in fact, the instructions made students aware that they will need to provide a significant number of details in order to enrich their speech. Teachers shared that they need to see what students really know and can do, so instruction must be cleared as well to achieve this outcome (Rudner & Schafer, 2002). Furthermore time also came into play, this point revealed valuable findings: The amount of items is proportional to the available time for the test. Surprisingly, there was no way to evaluate time since it was not stated in none of the exams.

A reminder should be on the exam for time since the oral exam is one of the productive skills evaluated the most and needs enough time. Finally, grading has an important role in testing: The test has a scoring/evaluation procedure that is specific and time-efficient. Teachers used a rubric and evaluated fluency, pronunciation, task

completion, vocabulary and language control on both exams. Also depicted the percentage that it will receive; however, there was no variety on the way of grading on the mid-term exam from the one on the final. There were teachers who admitted that if they took more time to carefully design a test, meaning items and organization, they would spend less time grading.

4. Writing Exam

(Mid-term)

This analysis was based only on the mid-term writing part, since the final exam did not have any writing part. In this analysis the theory stated on the theoretical framework will be used to and compare to the tests structure to judge how well testing principles have been applied or if there is something that can be modified or improved.

Principle of Validity

• Construct validity

The main objectives of the tests are not stated on it, but after reading the objectives on the syllabus, it can be seen that, as in the objectives says, students are allowed to express themselves by writing about atopic they are related to, as for example

the elections in the University of El Salvador, this gives students freedom to use the language to express their point of view. The way this part of the test was constructed includes topics of relevance for students making the writing more interest and less difficult. Nevertheless, teachers omitted a writing part in the final exam this leaves a gap on evaluation and prevent the tests to be constructed in a valid way since there is not structure to check for the final tests.

- **Content validity**

Content validity refers to tests what is taught in classes, here the writing part of the mid-term test was well structured because all the topics included in the choices students had to write about were related to students surrounding context. Furthermore, the topics were taught to give students the opportunity to write about something they really know about. To the researchers the topic selection was thoughtful and fair for students' level. In conclusion, as Fleurquin (2010) stated the aim objectives is to be oriented to students and this objectives must be reflected on the tests, this topics are clearly well selected to encourage students interest for writing.

- **Face validity**

The main feature of test face validity is that a test should look valid for the test taker (Ley,2007). In addition, students must be told, in advance, what is going to be evaluated, how it is going to be evaluated and time must be enough for the student to finish the test.

Comparing the writing part of the mid-term test to the theory presented on the theoretical framework the test is well organize and clear, the format and organization is pleasant for the reader eyes and the level of difficulty of the test is suitable for the students' level. Finally, the test includes clear instructions and criterions of evaluations for students to be aware of what they are going to be evaluated. In conclusion, this part of the tests was well designed taking into account all the face validity aspects, form tests appearance to test level.

Principle of Authenticity

The closer to reality a test is the more authentic it becomes (Lewkowicz, 200). Items and topics included on the test must be related not only to real situations but also to situations that students may encounter in their everyday-life. For this principle the test includes meaningful and authentic topics that are related to students' surrounded context; therefore, students have freedom to express themselves in the most natural way about topics they know about. In short, this part of the test reaches the standards of authenticity since it has been well designed to motivate students to write about real-life scenarios.

Principle of Practicality

Tests instructions are well designed and criteria of evaluation are included in the test to give students a clear idea as to how to write their paragraph. Besides, the time available for the test taker to finish it is enough and the topic facilitates the student to express freely. Although, this type of tests uses complementation items that are time consuming to grade the length of the paragraph is imitated to contain a topic sentence, some supporting details and at the end a concluding sentence. In conclusion, this test, is practical well designed to focus students on what they have to do and how.

5. Reading Exam

(Mid-term and Final)

This analysis will compare mid-term and final reading part. In order to see how well principles of testing are applied on this specific part on the tests administered on the Foreign Language Department, every principle will be discussed in order to establish how well it was applied.

Principle of Validity

- **Construct Validity**

As previous stated by McAlpine (2002) construct validity refers to the relation of what is being taught and how it is being evaluated, so that the way a topic is taught that same

way it should be evaluated. Both mid-term and final tests had two readings each, as stated in the syllabus objectives, these readings were about nowadays issues, but something very important stated in the objectives is that the readings should be related to the field and interest of students. However, one of the readings was not interesting for students' contexts or for students' field. In conclusion, teachers are trying to incorporate readings that are adjusted to students' context but in the mid-term exam they fail to fulfill their own reading skill objective, although teaching and evaluation are related.

- **Content Validity**

As the way of teaching has to be linked to the way of testing, the content taught has to be the same content evaluated (Kaplan & Saccuzzo, 2008). In order to accomplish reading content validity in a test, teachers must have in mind not only the contents taught in class but also the level of students. The reading tests both of them had readings with intermediate vocabulary suitable for students, but a topic seems to be too centralized on politics on the United States making the reading a little bit out of students' interest as stated on the objectives to this macro-skill. In short, good selections of readings, although these could be more focus to matters that students can see in their surrounding context.

- **Face Validity**

This principle regards to the viewpoint of the test taker as expressed Lay (2007). Test must seem valid for students. First of all the student must be used to the way of testing, if

students are not used to a certain kind of items this would add an extra level of difficulty to the test. For the reading part of the tests the format of the items were of multiple choice students on the focus group claim that this type of items were familiar for them in the reading part, so it can be said that this principle is achieved; however, some of the words evaluated were from technical language. Finally, the test format should be organized and neat, for the mid-term exam this principle was accomplished since the organizations of the test is pleasant to the eyes of the test taker. On the contrary, the final test looks awful, and is not pleasant to read.

Principle of Authenticity

Real-life use of the language must be present on a test in order to make it authentic (Igram, 2003). Students learn a language to use the language on their everyday life, but sometimes tests are not related to real life situations and students are evaluated something they may not use. In the reading part of the tests reading not only should be about nowadays matters but also about topics of interest for students. Although, the readings used in both tests are about nowadays situations some of them fail to catch students' interest and provide them with vocabulary they could use as soon as they go out of the classroom.

Principle of Practicality

This principle refers to the facility to score, interpret, design and administer the test. This part of the tests usually uses items that are easy to score and interpret most of the items were multiples-choice items, these types of items make the tests easy to score and the amount of items is not too much the length of the readings. In short this part of the tests, although is difficult to design to provide a real challenge, it is practical and easy to score and interpret.

CHAPTER 5

CONCLUSIONS

Testing is an important part of the learning process because it gives insights of students' learning and grades to report this progress. Testing is critical to decide whether a student passes or fails the course according to the syllabus evaluations system, so it is mandatory to give enough importance to tests design. With the triangulation analysis there were identified very important points linking the theory of testing principles with students' opinions and comparing this with the teachers' opinions. The researchers concluded the following.

Firstly, it was seen that one of main factors that affects students according to their opinion, as well as teachers' opinions, is the lack of adequate classrooms to take tests. The poor conditions inside the classroom with desks, room environment and equipment; as well as the conditions from the outside, make students' concentration lower and therefore grades are affected by this factor.

Besides, even though all teachers agreed to work in group, team work is not present since they just divide the test and there is no person in charge of checking the test to set the parts together. There is no specialized team to review exams and whether they comply with testing principles, therefore they are disregarded. This makes that tests do not really show what students have really learned from the course.

Furthermore, the research work proved that the test design part is not as relevant as the grading part is for some teachers. Some teachers admitted that due to the number of students on English courses the test design process is somehow expedite and that they are more worried on the grading part which in their opinion is the longest part. This in turn has a tremendous

impact on tests for teacher decision on what to include will be to put what is easier to grade rather than tasks that will give an insight on what students really can do with the language. Not only do teachers need time to grade tests but also time to evaluate and weigh what is going to be included on the tests.

It also was found through this research that feedback is given little importance in the testing process. In the focused group, when students were asked if they had feedback after exams some said that they had, while some others had not. Students were conscious on the importance of getting feedback so that they learn what the mistakes were to avoid making them in the future. Some teachers revealed that once the grading process had ended they checked if most failed then teachers went ahead explain the topic again and also made changes for the future. Feedback is also considered a cornerstone after judging students learning through a test; all teachers have to ask themselves what went wrong, what can be improved and what aspects they need to clarify; in other words, teachers need to give feedback otherwise the testing process is not complete.

It was discovered on exams that tests items were not contextualized for students' environment. Through this research it was found that some of the topics listed were not adapted to students' knowledge and daily life. This in turn could make students have a hard time when presenting since topics were not common. It seemed that exams were not double checked and that topics were chosen just because they enabled a discussion. There is nothing wrong on searching for items in the internet or any type of sources as long as teachers evaluate them before adding them to an exam. If teacher will be copying items from internet they should be as close as possible to students' knowledge.

According to students' opinions researchers found out that students are not aware about what they are expected to do at the end of the course; nonetheless, the syllabus established the main objectives of each language skill but neither tests nor teachers highlighted the importance of them during the course. Consequently, students do not know if they have reached those objectives. Besides, tests vaguely reflected the objectives that were supposed to reach when testing determined skills.

At last but not least, important researchers' finding was that the methodology used in class is not totally congruent with the testing process. The syllabus of the Intermediate Intensive English Class established that the methodology used in the course was going to be communicative, nevertheless, most of the procedures of testing lack communicative features. Therefore, paper and pencils tests were addressed to evaluate language skills such as listening, writing, grammar, reading and vocabulary for what students knew and not what they were able to do with the language.

RECOMMENDATIONS

After doing the investigation about testing application in the Foreign Language Department and taking into consideration teachers' and students' opinions, the researchers landed to the following recommendations which are intended to improve the testing experience of the English Students of the Foreign Language Department.

Even though it is not in the teachers' hands to change the classroom conditions, new ways to evaluate should to be promoted where the distractions such as noise are lesser. Although the Foreign Language Department's Language Laboratory is not completely well equipped, the students in the courses should be divided as to look for a way to examine students in small groups in the Lab, so avoiding to examine large groups in the classrooms where the noise is more present. With a good arrangement of teachers and students time, tests could be done in a better atmosphere.

When designing the tests, there should be more unity from teachers rather than just diving the parts of the tests. In the end, tests usually end being designed under teachers' beliefs. If done this, there should be a specialized team that checks whether each test, or the unique tests used for all the courses, complies with the testing principles so they become more reliable and valid to assess students. It is very important not to disregard these principles in order to be fair with students and with the lessons themselves to show what students have really learned and to seek improvement of teaching.

Testing is a very complex process; teachers are focusing on grades rather on test as the means to get those grades. Students may fail an exam for a variety of reasons: anxiety, fear, unprepared, and also because of test design. It is teachers concern to design the test in the best possible way and reduce any issue with the test design. Teachers should not worry only on grading

50 exams yet on applying all the testing principles so that the process of grading could be even easier. In this sense testing principles are very useful and should not be overlooked.

Providing students with feedback after revising tests should be mandatory. It is useless to hand in tests results without going over the aspects that made student get a low grade. Teachers need to remember that students can also learn from a test. If students clarify doubts, they will not make the same mistake in the future. Then, teachers' feedback is very valuable for students learning. Moreover, feedback is also important for teachers for they can highlight what aspects can be improved or changed in the way a test is being developed.

Any information gathered from Internet should be adjusted to student's knowledge when used on tests. Nowadays the Internet also provides information for educational fields yet these are made for a variety of purposes. It is the teachers' responsibility to review that information and adjust it as much as possible for their students. In this sense, teachers have to avoid copying and pasting because the information in the internet is very broad and teachers need to narrow them down by contextualizing the items.

A valid test is congruent with the methodology used in class and the methods of evaluation. Therefore, if communicative methodology is used in class teachers of the Foreign Language Department should use more tools of evaluation besides the oral exam in order to get to know what students are able to do with the language it means use communicative features in the tests administered to students.

Due to the importance of the objectives of the course and since they lead the learning of the second language, teachers should remind students the class objectives not only at the beginning of the course, but also during the semester. In this sense students will be aware about what they are expected to be able to do at the end of the course.

REFERENCES

- A description of educational measurement, assessment, and evaluation . 2012. A description of educational measurement, assessment, and evaluation . [ONLINE] Available at: http://www.adprima.com/measurement.htm. [Accessed 12 March 2012].

- Bachman, L.F. (2003). Building and supporting a case for test use. Language Assessment Quarterly, 2, 1-34.

- Bob Kizlik. (2012). Measurement, Assessment, and Evaluation in Education. Available: http://www.adprima.com/measurement.htm. Last accessed 31st Jan 2012.

- Bresciani, M. et al. (2009) A practical assessment, Research and Evaluation. http://pareonline.net, 14 (12), p.2.

- Brookes.ac.uk (2008) Oxford Centre for Staff and Learning Development. [online] Available at: http://www.brookes.ac.uk/services/ocsd/ [Accessed: 25 Mar 2012].

- Coombe, C. Folse. K. & Hubley, N. (2007). A practical guide to assessing English language learners. Ann Arbor, MI: University of Michigan Press

- David Nicol & Debra Macfarlane-Dick (2008) "Formative assessment and self-regulated learning: a model and seven principles of good feedback practice" Studies in Higher Education. Vol 31(2), 199-218.

- Doye, P. eds. (1991) Authenticity In Language Testing.

- Michael Milanovic. (2012) Principles of Good Practice, quality management and validation in language assessment. [e-book] University of Cambridge. Available through: cambridgeesol http://www.cambridgeesol.org/assets/pdf/general/pogp.pdf [Accessed: 11th September, 2011.].

- Educational Testing Service (2009). Guidelines for the Assessment of English Language Learners, Stanford 2005

- F. Fleurquin (2010). Test Development Stages and Specifications, EFL Assessment.

- Gabrenya, W. (2003) Validity. [e-book] http://my.fit.edu/~gabrenya/IntroMethods/eBook/validity.pdf [Accessed: 15Th Nobember, 2011].

- Grudens-Schuck, N. et al. (2012) Focus Group Fundamentals. [e-book] Iowa: Iowa State University of Science and Technology.

- Henry L. Roediger, III, Kathleen B. McDermott & McDaniel (2007). Using Testing to Improve Learning and Memory,

- Igram, D. (2003) Towards More Authenticity in Language Testing. [e-book] Brisbane: Griffith University.

- James H. McMillan, 2000. Essential Assessment Concepts for Teachers and Administrators (Experts In Assessment Series). 1 Edition. Corwin Press.

- Laerd Dissertation (2010) Convenience sampling: An overview. [online] Available at: http://dissertation.laerd.com/articles/convenience-sampling-an-overview.php [Accessed: 18th December, 2011.].http://www.extension.iastate.edu/publications/pm1969b.pdf [Accessed: 16TH November, 2011].

- Lado, R., 1961. Language Testing: The Construction and Use of Foreign Language Tests. Longman, London.

- Lyle F. Bachman, 1996. Language Testing in Practice: Designing and Developing Useful Language Tests (Oxford Applied Linguistics). Edition. Oxford University Press, USA.

- Mack, N. and Woodsong, C. (2005) Qualitative Research Methods: A data collector´s field guide, family health international, North Carolina.

- Marilla, Svinicki and Wilbert J., McKeachie. (2010). McKeachie's Teaching Tips: Strategies, Research, and Theory for College and University Teachers. Belmont, Wadsworth Cengage Learning.

- McAlpine, M. (2002). Principles of assessment. Glasgow: University of Glasgow, Robert Clark Center for Technological Education. Available at:

  http://www.caacentre.ac.uk/dldocs/Bluepaper1.pdf

- Morgan, D. L. (editor) 1993. Successful Focus Groups: Advancing the State of the Art. Newbury Park, Calif.: Sage

- Ontario. Ministry of Education. (2010). Assessment, Evaluation, and Reporting in Ontario School. Toronto: Author.

- Pedhazur, ElazarJ. and Lioras Pedhazur Schemelkin. Measurement, Design, and Analusis: An integrated Approach. Hillsdale, NJ: Lawrence Erlbaum Assosiates, 1991.

- Piontek, M. (2008) BEST PRACTICES FOR DESIGNING AND GRADING EXAMS. [e-book] michigan: Center for Research on Learning and Teaching. Available through: University of Michigan, CRTL http://www.crlt.umich.edu/publinks/CRLT_no24.pdf [Accessed: Novenber 15th, 2011].

- Philip Ley. "Quantitative Aspects of Psychological Assessment (2007)

- Professional Testing Inc. (2006) Test Validity. [online] Available at: http://www.proftesting.com/test_topics/pdfs/test_quality_validity.pdf [Accessed: 11th October, 2011.].

- Robert M., Kaplan and Dennis P., Saccuzzo. (2009). Psychological Testing Principles, Applications, and Issues. Pasadena: Wadsworth Pub Co.

- Rudner, L. and W. Schafer (2002) What Teachers Need to Know About Assessment. Washington, DC: National Education Association.

- Rust, C. (2002) Purposes and Principles of Assesment. [e-book] Oxford: Oxford Brookes university. Available through: www.brookes.ac.uk http://www.brookes.ac.uk/services/ocsd/2_learntch/briefing_papers/p_p_assessment.pd f [Accessed: Nobemver 15th, 2011].

- Scrivener , J (2005). Learning teaching: a guidebook for English language teachers, Macmillan books for teachers, Great Britain 2009.

- Spratt, M. et al. (2005) The TKT Course: Teaching Knowledge Test. 3rd ed. Cambridge: University of Cambridge.

- Susan, Armstrong and Sandii , Chan and Janne, Malfroy and Rosemary, Thomson. (Ed). (2008). Assessment Guide Implementing criteria and standards-based assessment. Sydney: University of Western Sydney.

- Traub, R. E. and Rowley, G. L. (1991), Understanding Reliability. Educational Measurement: Issues and Practice, 10: 37–45. doi: 10.1111/j.1745-3992.1991.tb00183.x

- Umbc.uoregon.edu (2009) E-Teacher Scholarship Program. [online] Available at: http://umbc.uoregon.edu/eteacher/ [Accessed: 25 Mar 2012].

- U.S. Department of Labor, Employment and Training administration (1999). Testing and Assessment: an Employer's Guide to Good Practices.

- Wiggins, G. P. (1993). Assessing student performance: Exploring the purpose and limits of testing. San Francisco: Jossey-Bass.