

**UNIVERSITY OF EL SALVADOR
SCHOOL OF ARTS AND SCIENCES
FOREIGN LANGUAGE DEPARTMENT**



TOPIC:

ENHANCING ENGLISH- LANGUAGE LEARNING THROUGH THE APPLICATION OF ASSESSMENT AND EVALUATION PRINCIPLES ON STUDENTS FROM THE FOREIGN LANGUAGES DEPARTMENT, FROM THE SCHOOL OF ARTS AND SCIENCES OF THE UNIVERSITY OF EL SALVADOR, DURING TERM II- 2020, AND TERMS I AND II- 2021

TÍTULO:

MEJORANDO EL APRENDIZAJE DEL IDIOMA INGLÉS A TRAVÉS DE LA APLICACIÓN DE PRINCIPIOS DE EVALUACIÓN EN LOS ESTUDIANTES DEL DEPARTAMENTO DE IDIOMAS EXTRANJEROS DE LA FACULTAD DE CIENCIAS Y HUMANIDADES DE LA UNIVERSIDAD DE EL SALVADOR, DURANTE EL CICLO II-2020 Y LOS CICLOS I Y II-2021

PRESENTED BY:

LUIS ANTONIO CHÁVEZ LAZO
ANGÉLICA BEATRIZ NÚÑEZ APARICIO
OSVALDO AARÓN HENRÍQUEZ RIVAS

CARNÉ

(CL14030)
(NA15010)
(HR14039)

IN ORDER TO OBTAIN THE DEGREE OF:

LICENCIADO/A EN IDIOMA INGLÉS: OPCIÓN ENSEÑANZA

DOCENTE ASESOR:

MAESTRO MAURICIO SALVADOR CONTRERAS CÁRCAMO

COORDINADOR GENERAL DEL PROCESO DE GRADO:

MAESTRO MIGUEL ANGEL CARRANZA CAMPOS

**CIUDAD UNIVERSITARIA, DR FABIO CASTILLO FIGUEROA,
SAN SALVADOR, EL SALVADOR, NOVEMBER 28TH, 2021.**

**AUTORIDADES DE LA UNIVERSIDAD DE EL
SALVADOR**

**RECTOR:
MAESTRO ROGER ARMANDO ARIAS ALVARADO**

**VICERRECTOR ACADÈMICO:
DOCTOR. RAÚL ERNESTO AZCÚNAGA LÓPEZ**

**VICERRECTOR ADMINISTRATIVO:
INGENIERO JUAN ROSA QUINTANILLA**

**SECRETARIO GENERAL:
INGENIERO FRANCISCO ANTONIO ALARCÓN SANDOVAL**

**FISCAL GENERAL:
LICENCIADO RAFAEL HUMBERTO PEÑA MARÍN**

**AUTORIDADES DE LA FACULTAD DE CIENCIAS Y
HUMANIDADES.**

**DECANO:
MAESTRO ÓSCAR WUILMAN HERRERA RAMOS**

**VICEDECANA:
MAESTRA SANDRA LORENA BENAVIDES DE SERRANO**

**SECRETARIO:
MAESTRO JUAN CARLOS CRUZ CUBIAS.**

TABLE OF CONTENTS

ABSTRACT	1
INTRODUCTION.....	2
1. RESEARCH QUESTIONS	3
MAIN RESEARCH QUESTION	3
SUBSIDIARY QUESTIONS.....	3
2. RESEARCH OBJECTIVES.....	4
2.1. GENERAL OBJECTIVE.....	4
2.2. SPECIFIC OBJECTIVES	4
3. IMPORTANCE	5
4. THE METHOD	6
4.1. Type of Research: non-experimental/ observational study	6
4.2. Study Design	6
4.3. Study Population.....	7
4.4. Sample.....	7
4.5. Sample selection process	8
4.6. Techniques and Research Instruments	8
4.6.1. Research Technique: Bibliographical Review	8
4.6.2. Research Instrument: bibliography card	9
4.6.3. Research Instrument: Checklist	10
5. FINDINGS.....	11
5.1. The Importance of the Principles for Assessment and Evaluation	11
5.2. Testing and Assessment.....	12
5.3. Principles of Assessment.....	15
5.3.1. Validity and Reliability	15
5.3.1.1. Types of Validity and Reliability	16
5.3.1.1.1. Content Validity	17
5.3.1.1.2. Construct validity	17
5.3.1.1.3. Face Validity	18
5.3.1.1.4. Criterion Validity	18
5.3.2. Reliability	19
5.3.2.1. Administration Reliability	20
5.3.2.2. Student-related Reliability	20

5.3.2.3. Rater Reliability	21
5.4. Authenticity	21
5.5. Practicality	22
5.6. Usefulness	22
5.7. Washback	23
6. THE EXPECTED RESULTS	23
7. CONCLUSIONS	29
8. REFERENCES.....	31
9. APPENDICES.....	33
9.1. Bibliography Cards.....	34

ABSTRACT

This research intends to know how important is to clarify some of the fundamental principles and issues which need to be applied to the design of any assessment strategies for any module or programed. For this purpose, let us define assessment as evaluation or appraisal; it is about making a judgment, identifying the strengths and weaknesses, the good and the bad, and the right and the wrong in some cases. It is more than simply giving marks or grades, although that may well be a part of it. And because it involves making a judgment it will almost inevitably include an element of subjectivity by the assessor. However, we should strive to make assessment as objective, fair and transparent as possible. Assessment plays a crucial role in the education process: it determines much of the work students undertake (possibly all in the case of the most strategic student), affects their approach to learning and, it can be argued, is an indication of which aspects of the course are valued most highly.

Key Words: Assessment, reliability, validity, authenticity, construct, rubric, Evaluation.

INTRODUCTION

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 they do not fulfill the principles of test design: reliability, validity, authenticity, practicality, washback, and usefulness. Making up good tests guarantees good 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 they actually comply with all the principles previously mentioned.

In the Foreign Languages Department, from the School of Arts and Sciences of the University of El Salvador, the importance of the quality of the tests administered to students has been questioned, but has never been investigated from a theoretical perspective. Therefore, this research attempts to explore the literature on the application of testing principles on test design and administration so that this bulk of theoretical analysis can be considered in the evaluation and assessment endeavor by teachers in the English field of the Foreign Language Department of the University of El Salvador. The shortage of research in this area makes this study more significant since it would be an antecedent and a reference for future research projects.

1. RESEARCH QUESTIONS

MAIN RESEARCH QUESTION

- How does the application of assessment and evaluation principles on test design and administration improve the quality of the teaching and learning processes of students from the FLD of the School of Arts and Sciences from the University of El Salvador?

SUBSIDIARY QUESTIONS

- How does the application of assessment and evaluation principles improve test design and administration to students from the Foreign Language Department from the University of El Salvador?
- What are the effects of not applying the testing principles on exams when evaluating students from the Foreign Language Department from the University of El Salvador?
- What are some recommendations for the application of assessment and evaluation testing principles in exams administered at the Foreign Language Department from the University of El Salvador?

2. RESEARCH OBJECTIVES

2.1. GENERAL OBJECTIVE

- To explore how the application of the assessment and evaluation principles improves the quality of teaching and learning English as a foreign language in teachers and students from the Foreign Languages Department from the University of El Salvador.

2.2. SPECIFIC OBJECTIVES

- To describe the process of applying the assessment and evaluation principles in test design and administration in exams administered in the Foreign Language Department from the University of El Salvador.
- To determine to what extent not applying the theory on the application of assessment and evaluation principles affects both, the teaching and learning processes in the Department of Foreign Languages from the University of El Salvador.
- To offer recommendations for the application of assessment and evaluation testing principles in the exams administered at the Foreign Language Department of the University of El Salvador.

3. IMPORTANCE

Testing is used to gather information on what students know and can do; As such, it needs to be subject to scrutiny on how it is designed and what is included. Judgements from teachers, based on responses from students are very important aspects to be considered, when a test is created, to conclude whether students are progressing or not. Such aspects have to do with testing principles guidance. For instance, considering reliability and validity is a good way of weighing 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. (Kaplan & Saccuzzo, 2008).

There is not much information about applying testing principles for testing in the Foreign Language Department, from a deep theoretical perspective. Indeed, research projects on these testing issues have not been carried out with students on the English Developmental Areas in the Department. Due to this fact, and since tests have become a major area of study within educational research, this study tried to find out more information on this topic, focusing its efforts on the Foreign Language Department, and the English Teaching in this academic unit. In short, in the light of the shortage of knowledge, this research aims to contribute to fill up this gap by gathering valuable information on the application of testing principles to improve tests quality.

4. THE METHOD

4.1. Type of Research: non-experimental/ observational study

Observational research: Observational studies are ones where researchers observe the effect of a risk factor, diagnostic test, treatment or other intervention without trying to change who is or isn't exposed to it (Institute for Work and Health, 2016). In other words, this type of study aims to present the reality of a fact, regarding the two variables (factors) being considered for the study. Just as in this study, the researchers did not try to interfere with these phenomena, but they observed them in order to explain whatever they could see about it. As previously stated, the research group tried to explain the effect of one of the independent variables onto variable one.

The category of non-experimental designs is the most heterogeneous of the three classification categories (experimental, quasi-experimental, and non-experimental). Although, in general, this category has the lowest level, each design within this category varies as to its own individual level of scientific validity. Commonly, non-experimental studies are purely observational, and the results are intended to be purely descriptive. (U.S. Department of Labor, 1999).

4.2. Study Design

Qualitative research: As for this study, the researchers are using the qualitative research design. As stated by Bhandari (2020), "qualitative research involves collecting and analyzing non-numerical data to understand concepts, opinions, or experiences. It

can be used to gather in-depth insights into a problem or generate new ideas for research.” Regarding this definition, the group presents results that are useful for describing and explaining the usefulness of the application of the Evaluation and Assessment principles to the learning process of students of the Foreign Language Department from the University of El Salvador. But all this information is presented in the form of non-numerical data, through a glance at previous studies on this field (in spite of their shortage), as well as through a sample of tests that were designed at this academic unit.

4.3. Study Population

For this research, the group collected a package of 25 bibliographical resources, including materials such as books, websites, tests from the FLD, and previous research works. An important factor to include them was that they were somehow related to the research topic. These materials were subdivided as follows:

- Books: 5 books
- Websites: 10 websites (from trustable sources)
- Tests from the FLD: 3 tests
- Previous research works: 7 research works

4.4. Sample

After having selected the materials listed above, the group undertook a rigorous process to decide which of them should be used for this study, and which ones could be left apart. There were some criteria that the investigators used to make that decision, among such factors, and probably the most important one, there was the extent to

which each material was related to the research topic for this study. The most related materials were selected for this study, and the ones that had less relation were left apart. In the end, the group chose 6 bibliographical references for this study.

4.5. Sample selection process

Handpicked sampling: According to Zina O’Leary (2004), “handpicked sampling involves the selection of a sample with a particular purpose in mind, selecting cases that meet particular criteria.” As mentioned before, in order to select the materials that best fit this study, the group had to examine them by means of some criteria, so that only those that had major relevance for the study were included. In other words, the materials previously gathered had to go through some evaluation aspects in order to be chosen for this study. Finally, the 6 more appropriate materials were taken into account for this study.

4.6. Techniques and Research Instruments

4.6.1. Research Technique: Bibliographical Review

The analysis of the information started by choosing the correct sources. This means that the researchers narrowed down the search by taking into account the following theoretical constructs: Assessment and Evaluation Principles, and English-language learning. As such, this bibliographic research gathers information from previously published materials. These materials include more traditional resources such as books, magazines, journals, newspapers, and reports; it also contains more modern materials such as websites, blogs, and bibliographic databases. The convenience of online

information sources is an advantage; however, the speed of accessing material should not outweigh the need for quality or reliability of the content; that's why the group rather carefully chose each material instead of doing it at a fast speed. All in all, the challenge for the researchers was discerning which resources were appropriate to include in academic research and how to do the best to enrich this bibliographical research.

Guillermina Baena (1985) stated that "documentary (bibliographic) research is a technique that consists in the selection and collection of information through reading, critique of documents and bibliographic materials, libraries, newspaper archives and educational centers. documentation and information." In other words, a bibliographic study is based on previous materials about the field being investigated. This definition certainly applies to the type of research being conducted.

4.6.2. Research Instrument: bibliography card

When utilizing bibliography cards, researchers should follow some stages such as:

- a. Give proper credit to all sources to avoid any hint of plagiarism.
- b. Write down all necessary documentation information before taking notes.
- c. Use one index card for each source and use a special color or size for bibliography cards.
- d. Consult the Bibliography/Works Cited page to be sure you list all the necessary information and punctuate it correctly.
- e. Write neatly.

(Taken from Springfield Public Schools, 2015)

Based on this information, a bibliography card can be defined as an instrument used to gather the information about the sources being used to support a study, providing it with background information for it, in which the investigators also give the appropriate credits to the author(s). In this way, the researchers present a list of the materials that the author is using so as to support his research.

4.6.3. Research Instrument: Checklist

“A checklist is a scoring tool that explicitly represents the performance expectations for an assignment or piece of work.” (Carnegie Mellon Center, 2021). Keeping this definition in mind, the researchers decided to use a checklist in order to select the sources (previous literature) that were used for this study. This instrument allowed the group to make an objective decision on which of the sources gotten for the population should be kept as the sample. The ones that got higher scoring were selected for this research in the end.

5. FINDINGS

5.1. The Importance of the Principles for Assessment and Evaluation

Not all teachers know the importance of the assessment and evaluation principles; however, their importance should be taken into consideration in the field of education when designing instruments for evaluation and assessment, for this aspect measures the learning outcomes through accurate conclusions based on the test results by students.

Another aspect that is extremely important when making tests is reliability. This principle is used to assign grades or marks to students' learning. Then, there is the authenticity principle, which refers to how authentic the language of a test is. To illustrate this, let's figure out that the examinees will read some piece of news from a newspaper. Since this is a piece of text made to communicate real facts in a real means of communication, the language that they will come across will be natural; it means, the very kind they will find and use when communicating. On the other hand, when the material is created to specifically present or practice a particular structure, language may become artificial, meaning it may differ from the type used in real communication. All in all, and according to the principle of validity, it is better to face students with the kind of language they will find in the real world, even for exams.

Practicality is also an important principle for the process of evaluation since teachers need not only to be aware of what will be placed on the design but also to be aware of the means the teacher will use to design a test. In addition, the usefulness of a test mainly depends on which purpose it is intended to fulfill as well as how aware students

are of this reality. Another important element to regard when designing a test is washback. It is an important fact that tests have an effect on students' learning; they are powerful determiners of what happens in the classroom in the teaching- learning process.

Along this research paper, these principles will be explained in more detail. The criteria and the gathering testing principles are vitally important in assessment. In fact, testing principles can be of great help for teaching and research. It is important to apply the principles for assessment and evaluation when testing students so that both the learning and teaching experiences can be improved in the Foreign Language Department of the School of Arts and Sciences of the University of El Salvador.

There, the importance of the quality of tests administered to students has been questioned but has never been investigated from a theoretical perspective. Therefore, this research attempts to explore the literature on the application of testing principles on test design and administration, so that this bulk of theoretical analysis can be considered in the evaluation and assessment endeavor by teachers in the English field. The shortage of research in this area makes this study even more significant since it would be an antecedent and a reference for future research projects on this area.

5.2. 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. Likewise, in some countries, people must 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.

Now, it is important to clarify some concepts. These concepts are assessment, evaluation, and testing; also, there is a need to explain how different they are. To start with, assessment is an academic process that 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 defined. But what is the difference between assessment and evaluation? For this research, evaluation will be conceptualized as "the process of judging the students' learning based 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 based on a certain criteria or standards as level of grammar use, spelling, fluency, and accuracy. These judgments are what is called evaluation.

In the field of education, testing is especially important to provide valuable information on 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 the one hand, assessment is any activity conducted to know about students’ learning progress. 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 of 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 they have 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). Through 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 to know 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 about the learning process, it is important to consider certain assessment 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.

5.3. Principles of Assessment

There is a variety of issues that must be kept in mind 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 should 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 research work explains and clarifies the definition of each principle of assessment, so that teachers can know how to apply them in their assessment process.

5.3.1. Validity and Reliability

In spite of the fact that validity and reliability are two different things, McAlpine (2002) joins and defines them as the overall quality of assessment. The relationship between them is that a valid test must always be reliable. This means that teachers not only need to seek for reliability but also for validity in a test. In order to decide if a test is valid and reliable, both must be present. To define more deeply these principles, it can be said that validity refers to the capacity to measure what it 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).

5.3.1.1. Types of Validity and Reliability

The University of Oregon (2005) has taken a close look into subdivisions in regards to validity and reliability. Talking about test validity, it can be said that content validity exists if what is being addressed and included in the test has been explained during instruction, enhancing student's 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).

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 a test measures. All the items in a test need to be related

with the purpose of the test. They should also reflect what the students expect from the tests. The following are the aspects that should be taken into account when creating a test.

5.3.1.1.1. 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 no content validity.

5.3.1.1.2. 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).

5.3.1.1.3. 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 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. If the test measures what students learnt but in a completely different way from the one it was taught, there is no face validity because the test takers will feel that they are different from what they were taught.

5.3.1.1.4. 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 criteria (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 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.

5.3.2. Reliability

In education, tests are used to assign grades or marks to the students' learning, but the accuracy of these grades or marks is not always satisfactory to the student, parents, or teachers. Tests always have certain discrepancies between students' real ability and the measurement of the ability (Kaplan & Saccuzzo, 2008). These discrepancies are seen as errors of measurement, but this does not mean that tests are useless to measure students' ability. This is just inaccuracy that measurements have. Due to the existence of these errors, teachers need to look for ways to minimize them on their tests. This notion leads to the concept of reliability. What is reliability on tests, then? 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 by some factors, which are included as part of reliability achievement, as follows:

5.3.2.1. Administration Reliability

The term “administration reliability” refers to the setting or the conditions under which 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. Besides, instructions given to students have to be considered. Teachers must be careful when explaining or giving instructions to the examinees in order to avoid that they can guess the answers. The last, but not least, important condition is time limits. Teachers should consider if the time is sufficiently 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.

5.3.2.2. Student-related Reliability

Regarding the subdivisions of reliability, student-related reliability is found. This factor takes into consideration emotional aspects like those that, at the 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 factors, anxiety, temporary illnesses, and fatigue can be also found. In fact, the influence of feelings such as 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, when a person is prepared and best able to perform well, also ensures that students

understand what is expected and will avoid failing this principle (Coombe & Hubley; 2007).

5.3.2.3. Rater Reliability

Reliability is an important asset in education. It is specifically important in testing, and it is shown when results 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 is more carefully assessed.

5.4. 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 items, 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 focused on isolated grammar patterns or language knowledge (Ingram, 2003).

5.5. Practicality

In addition, practicality is 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 the teacher will use to do such a test. Then, a question arises and asks if our assessment meets our learners' needs within available resources. By this, it 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).

5.6. Usefulness

The usefulness of a test mainly depends on which purpose it is intended to fulfill as well as how aware students are of this reality and needs to get 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.

5.7. Washback

One important fact is that tests also influence 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.

6. THE EXPECTED RESULTS

In El Salvador, it is well-known that getting a degree at the University of El Salvador is not a piece of cake. Students from the Foreign Language Department face several situations that make these years of study even more difficult than what it was supposed to be. Resources, infrastructure, and teachers' behaviors are only a few things on the list of things students are faced with. During the development of this work and taking into account what the researchers have written about the topic, it cannot be denied that the principles of assessment could solve some of the issues that are found at the Foreign Language Department if they were applied properly by the authorities and the teachers. How the application of those principles could enhance English language learning at the Foreign Language Department, based on *Principles of Assessment* by *Mhairi McAlpine 2002*, will be presented below:

- 1- Even though it is not in the teachers' scope to change the classroom conditions, new ways to evaluate should be promoted, where the distractions such as noise are less. Students often take exams in which it is impossible to listen to the audios played in those big classrooms with 40 students and a lot of noise outside (This happens when classes used to be in the face-to-face modality). Although the language laboratory in the Foreign Language Department is not completely well equipped, the students in the courses should be divided so as to look for a way to examine them in small groups in there; in this way, it is easy to avoid the examination of large groups in classrooms where the noise is louder. With a good arrangement of teachers' and students' time, tests could be carried out in a better atmosphere.

2- When designing the tests, there should be more unity from teachers rather than just dividing the parts of the tests. In the end, tests usually end up being designed under teachers' beliefs. There must be a special team of teachers that check whether tests are appropriate for the students' English level, taking into account what students have studied so far to determine if those tests comply with the testing principles, so they become more reliable and valid to assess students. Students could take exams in which structures and topics that had not been taught in classes were covered, which affected their grades. That is why it is very important not to disregard these principles to be fair with students and the lessons themselves to show what students have really learned. Checking that all the exams made by teachers meet the criteria related to all the assessment principles is a must; in that way, teachers can avoid exams that are not well-prepared, and which can jeopardize student's grades and learning.

3- Testing is a very complex process. Sadly, some teachers focus on grades rather than on tests as the means to get those grades. Students may fail an exam for a variety of reasons: anxiety, fear, being unprepared, and because of a bad test design. There are students who feel afraid of some teachers because they have a bad reputation for designing really difficult exams. It was also seen that those teachers feel somehow proud of that reputation instead of worrying about the student's knowledge. It is the teacher's responsibility to design tests the best possible way to reduce any issues with test design. Teachers should avoid creating exams that only seek to make grading an easier job. In this way, testing principles are very useful and should not be overlooked. Prior explanation

to students regarding the structure of the exams can help them reduce the anxiety they feel before an exam; being so, they can be prepared for what the test will look like, and they will only need to focus on the content to be reviewed to perform well in that test. The Foreign Language Department authorities should enhance teachers to stop surprising students with test structures that the only thing they cause is confusion to students.

4- Providing students with feedback after revising tests should be mandatory. It is useless to give test results without going over the aspects that made students get a low grade. Teachers need to remember that students can also learn from mistakes and feedback. If students clarify doubts, they will not make the same mistake in the future. Then, feedback from the teachers is valuable for students' learning. Moreover, feedback is also important for teachers, so they can highlight what aspects can be improved or changed in the way a test is being developed and designed. At the Foreign Language Department, it is not common that teachers explain exams after they are graded, and students do not get an answer to the doubts they have or the reason why they made a mistake. Professors say that students learn from mistakes, but in this case, students do not learn since they do not know what the mistakes were. Teachers must give feedback after exams, for this will be meaningful for students' learning.

5- Any information gathered from the internet should be adjusted to the students' 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 amount of information on the internet is huge, and teachers need to narrow it down by contextualizing the items. Students often face this situation while working on the practice exercises; students need to collect ideas for the topics being studied by that time, and the same activity could be found on the internet, from an online course. Nowadays, thousands of exams can be found on the internet; it is easy to print them and give them to students, but this is not really helpful for them since the content in those exams might not be the appropriate for the English level of the students, and this will only cause them to get low grades and a low self-esteem level due to a content that was not supposed to be considered to evaluate them at that moment. This must be avoided at the Foreign Language Department; instead, teachers need to get together and plan their exams taking into account the content studied up to that day.

6- Because of the importance of the objectives of the course and since they lead the learning of the language, teachers should remind students of the class objectives not only at the beginning of the course, but also during the semester. In this way, students will be aware of what they are expected to be able to do at the end of the course. Students could be motivated at the beginning of the class, after a couple of days they can forget the objectives if teachers do not take a couple of minutes to discuss as a class if they had achieved those objectives. Teachers from the FLD need to emphasize on those objectives along the whole

course. That way, students can be encouraged and motivated to achieve those goals at the end of the semester.

Principles of Assessment by Mhairi McAlpine February 2002

7. CONCLUSIONS

As researchers, the group was able to learn and discover the correct form of applying the 7 principles to elaborate tests that will fulfill the needs not only for teachers but also for students. (*Principles of Assessment by Mhairi McAlpine, February 2002*) Following the information presented in this research paper, the team can prove that testing is an important part of the learning process because it gives insights into students' learning and grades to report this progress.

It was seen that one of the main factors that affects students, according to their opinion, as well as teacher's opinions, is the lack of adequate classrooms to take tests. Besides, even though all teachers agreed they work in groups, teamwork 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.

Furthermore, this research work proved that the test design task is not as relevant as the grading part, for some teachers. Some of them admitted that due to the large number of students in the English class groups, the best design process is somehow expediting, and that they are more worried about the grading part, which, in their opinion, is the longest and the most difficult part.

The 7 principles for assessment help discover that test items were not contextualized for student's environment on exams. 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. This means using communicative features in the tests administered to students.

Last, but not least, important researchers found that the methodology used in class is not totally congruent with the testing process. The principles for assessing definitely help improving the learning process in the student, but it is necessary to follow the structure presented before.

8. 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].
- Bob Kizlik. (2012). Measurement, Assessment, and Evaluation in Education. Available: <http://www.adprima.com/measurement.htm>. Last accessed 31st Jan 2012.
- Philip Ley. "Quantitative Aspects of Psychological Assessment (2007)
- 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: November 15th, 2011].
- 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.
- Rust, C. (2002) Purposes and Principles of Assessment. [e-book] Oxford: Oxford Brookes university. Available through: www.brookes.ac.uk
- 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.
- Wiggins, G. P. (1993). *Assessing student performance: Exploring the purpose and limits of testing*. San Francisco: Jossey-Bass.

9. APPENDICES

9.1. CHECKLIST TO EVALUATE BIBLIOGRAPHICAL RESEARCH WORK	
ELEMENTS	COMMENTS
1. Statement of the problem	The problem or statement of the problem that has been found during the bibliographical work is relevant and describes a delimited and concrete problem which encourages you to investigate enthusiastically in order to find positive results.
2. Organization	The information has a good structure since it is well organized and coherent. The ideas are presented in a logical order. The comprehension of concepts and theories is easy to understand and interpret. The transition of ideas is also fluent which helps with the finding of positive results based on the statement of the problem.
3. Writing	There are no writing errors, spelling, and punctuation as well as the structure of the information is correct.
4. Quality of the information.	The information that the group has gathered to research the flipped classroom methodology is very well related with the topic. It gives clear ideas, as well as concepts, examples, main and secondary ideas, implementation, advantages, and disadvantages, etc. which has helped the group finding the information needed for the research on the topic flipped classroom.
5. Materials and Procedures	The different works found describe the material used during the research through their corresponding references. Also, they present in very detailed way how the process of the research during every step.
6. Conclusion	The conclusion is clear and specific which gives answer to the statement of the problem as well as the questions proposed at the beginning of the research.

9.1. Bibliography Cards

Bibliography Cards

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



UNDERGRADUATE PROJECT

Bibliography card

Topic: “ENHANCING ENGLISH-LANGUAGE LEARNING THROUGH THE APPLICATION OF ASSESSMENT AND EVALUATION PRINCIPLES ON STUDENTS FROM THE FOREIGN LANGUAGES DEPARTMENT, FROM THE SCHOOL OF ARTS AND SCIENCES OF THE UNIVERSITY OF EL SALVADOR, DURING TERM II- 2020, AND TERMS I AND II- 2021”

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

Source 1

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: November 15th, 2011].

Bibliography Cards

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



UNDERGRADUATE PROJECT

Bibliography card

Topic: “ENHANCING ENGLISH-LANGUAGE LEARNING THROUGH THE APPLICATION OF ASSESSMENT AND EVALUATION PRINCIPLES ON STUDENTS FROM THE FOREIGN LANGUAGES DEPARTMENT, FROM THE SCHOOL OF ARTS AND SCIENCES OF THE UNIVERSITY OF EL SALVADOR, DURING TERM II- 2020, AND TERMS I AND II- 2021”

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

Source 2

Philip Ley. (2007) “Quantitative Aspects of Psychological Assessment.

Bibliography Cards

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



UNDERGRADUATE PROJECT

Bibliography card

Topic: “ENHANCING ENGLISH-LANGUAGE LEARNING THROUGH THE APPLICATION OF ASSESSMENT AND EVALUATION PRINCIPLES ON STUDENTS FROM THE FOREIGN LANGUAGES DEPARTMENT, FROM THE SCHOOL OF ARTS AND SCIENCES OF THE UNIVERSITY OF EL SALVADOR, DURING TERM II- 2020, AND TERMS I AND II- 2021”

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

Source 2

Professional Testing Inc. (2006) Test Validity. [online] Available at:

http://www.proftesting.com/test_topics/pdfs/test_quality_validity.pdf [Accessed: 11th October, 2011.].

Bibliography Cards

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



UNDERGRADUATE PROJECT

Bibliography card

Topic: “ENHANCING ENGLISH-LANGUAGE LEARNING THROUGH THE APPLICATION OF ASSESSMENT AND EVALUATION PRINCIPLES ON STUDENTS FROM THE FOREIGN LANGUAGES DEPARTMENT, FROM THE SCHOOL OF ARTS AND SCIENCES OF THE UNIVERSITY OF EL SALVADOR, DURING TERM II- 2020, AND TERMS I AND II- 2021”

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

Source 3

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

Bibliography Cards

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



UNDERGRADUATE PROJECT

Bibliography card

Topic: “ENHANCING ENGLISH-LANGUAGE LEARNING THROUGH THE APPLICATION OF ASSESSMENT AND EVALUATION PRINCIPLES ON STUDENTS FROM THE FOREIGN LANGUAGES DEPARTMENT, FROM THE SCHOOL OF ARTS AND SCIENCES OF THE UNIVERSITY OF EL SALVADOR, DURING TERM II- 2020, AND TERMS I AND II- 2021”

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

Source 4

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

Bibliography Cards

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



UNDERGRADUATE PROJECT

Bibliography card

Topic: “ENHANCING ENGLISH-LANGUAGE LEARNING THROUGH THE APPLICATION OF ASSESSMENT AND EVALUATION PRINCIPLES ON STUDENTS FROM THE FOREIGN LANGUAGES DEPARTMENT, FROM THE SCHOOL OF ARTS AND SCIENCES OF THE UNIVERSITY OF EL SALVADOR, DURING TERM II- 2020, AND TERMS I AND II- 2021”

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

Source 5

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

Bibliography Cards

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



UNDERGRADUATE PROJECT

Bibliography card

Topic: “ENHANCING ENGLISH-LANGUAGE LEARNING THROUGH THE APPLICATION OF ASSESSMENT AND EVALUATION PRINCIPLES ON STUDENTS FROM THE FOREIGN LANGUAGES DEPARTMENT, FROM THE SCHOOL OF ARTS AND SCIENCES OF THE UNIVERSITY OF EL SALVADOR, DURING TERM II- 2020, AND TERMS I AND II- 2021”

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

Source 6

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