

BA Students' Perceptions of Standardized English Tests*

Percepciones de los estudiantes de licenciatura sobre las pruebas estandarizadas de inglés

* Original Research Article

Enrique Vez López**

Karla Elizabeth Molina Muñoz***

Diana Iveth Sánchez Hernández****


Reception: April 01, 2022


Approval: June 15, 2022


 <https://doi.org/10.19053/2011835X.14752>

How to cite this article:

Vez, E., Molina, K.E. & Sánchez, D.I. (2022). BA Students' Perceptions of Standardized English Tests. *ENLETAWA Journal*, 15(1), 1-31.

** Doctor en Ciencias del Lenguaje, Universidad Veracruzana. Docente de la Universidad Veracruzana, México. vez@uv.mx  <https://orcid.org/0000-0003-1117-5386>.

*** Maestra en Enseñanza del inglés como lengua extranjera, Universidad Veracruzana. Docente en el sector privado, México. elizabeth25750@gmail.com  <https://orcid.org/0000-0002-9006-3974>.

**** Maestra en Enseñanza del inglés como lengua extranjera, Universidad Veracruzana. Docente de la Universidad Veracruzana, México. diasanchez@uv.mx  <https://orcid.org/0000-0002-1426-5814>



Abstract

This research examines the perceptions of a group of BA students of English as a foreign language regarding standardized English tests at a school of languages in the state of Veracruz, Mexico. In this school program, all students take the same evaluations in order to assess their English proficiency level. This research study made use of a questionnaire containing both open-ended and closed-ended questions to examine students' viewpoints about tests design, validity, fairness and reliability. Numeric data as well as the participants' viewpoints and comments were used to support and present the findings. By analyzing the students' questionnaire answers, it was possible to identify whether the participants thought this type of assessment was fair, reliable, and valid in evaluating their proficiency level. In addition, this research helped identify weaknesses and strengths of standardized exams in this context.

Keywords: standardized exams, assessment, EFL, students' perceptions.

Resumen

Este reporte de investigación examina las percepciones de un grupo de estudiantes de inglés como lengua extranjera con respecto a los exámenes estandarizados aplicados en una facultad de lenguas en el estado de Veracruz, México. En este programa académico, todos los estudiantes toman los mismos exámenes, cuyo fin es evaluar su nivel de competencia en inglés. Para este estudio se utilizó un cuestionario con preguntas abiertas y cerradas con el objetivo de examinar los puntos de vista de los participantes en lo que se refiere al diseño de los exámenes, su validez, e imparcialidad. En el presente reporte, se utilizan tanto datos numéricos como las opiniones de los estudiantes para apoyar la presentación de resultados. Al analizar las respuestas al cuestionario fue posible identificar si los participantes consideraban que este tipo de exámenes son justos, confiables, y válidos para evaluar su competencia en inglés. Adicionalmente, esta investigación ayudó a identificar debilidades y fortalezas de los exámenes estandarizados en este contexto.

Palabras clave: Exámenes estandarizados, evaluación, inglés como lengua extranjera, percepciones de los estudiantes.

Introduction

Evaluation is an essential part of the learning process. There are different forms of assessing students; nevertheless, standardization has become one of the most used. Standardized tests “are used to assess a student’s mastery of the curriculum and to evaluate teacher and school effectiveness” (Graham & Neu, 2004, p. 296) When evaluating students’ progress, the main objective is to discover if they have a successful learning process. Owing to this, standardization appears to be the best option because it focuses on evaluating effectiveness.

This type of evaluations “assure that scores obtained by different people are comparable to one another” (U.S. Congress, Office of Technology Assessment, 1992, p.166). Furthermore, the U.S Congress states that another advantage of standardized tests is that all students are given the same amount of time, instructions and materials reducing external variables on the students’ score. In other words, standardization gives students the same opportunity to prove their knowledge on a specific subject or topic.

As stated before, it is believed that standardized tests are the best solution for assessing students; however, not all students are the same. As Taylor (2011) states, standardized tests are unlikely to show the students’ overall knowledge because sometimes what they assess is not an accurate reflection of students’ abilities. It might be that a student is good at producing the language but he or she does not show the same proficiency when being evaluated.

The Department for Education in the United Kingdom (2011) states that the increasing use of standardized tests has led to concerns about whether they effectively serve their purpose. The purpose of these evaluations is to measure the knowledge of pupils. However, as Taylor (2013) argues, pupils have different skills, abilities, and even the gifted students may feel pressure due to the tests, which might affect their academic success. Taylor (2013) goes on to state that assessing all students with an inflexible test is not the best option.

As aforementioned, standardized tests are used in different subjects and English is a field in which these tests have an important role. The Department for Education in the United Kingdom (2011, as cited in Taylor, 2013, p. 4) states that “standardized testing (ST) has become an instrumental part of the English educational context for over the last three decades”. Proficiency English tests are a clear example of the use of standardization. The purpose of this, as Benavides (2015) states, is to evaluate the use of the language in non-native people, focusing on vocabulary, grammar, listening comprehension, and writing.

The fact that students answer the same test may seem to be the best option to measure their proficiency level of English. However, it is important for tests to be well designed in order to measure students' proficiency correctly. When deciding to use standardization to evaluate English as a foreign language, it is important to consider how students perceive this type of evaluations to know if they view them as a reliable form of evaluation and to discover whether they really are serving their purpose.

In Mexico, these ST are used to evaluate different subjects; assessing English as a foreign language is an example. This study focuses on a small group of EFL students at the BA level in order to know their perceptions about the use of standardized English tests to measure their proficiency in English. The purpose of standardization in this context is to evaluate students' performance and gains in English courses. It is necessary to know what students think about this type of evaluation and whether they trust them to measure their proficiency.

It must be highlighted that there is a lack of research about the students' perceptions of these tests in the context of this research. Apparently, there does not seem to be much concern about standardized English tests or what the students think of them. Students' views are important because they might help us to understand if this type of assessment is the best option for that specific purpose.

The main objective of these tests is to measure the level of knowledge pupils have acquired during their lessons. Therefore, it is important to discover to what extent

students think these tests are fair and whether they trust them to evaluate their command on English. It is critical to assess if the use of standardized English tests is positive for this particular context.

Students' views about these tests as well as their thoughts and points of view are crucial when it comes to analyzing the efficiency of these evaluations. On the other hand, knowing what students think about these tests might help to improve or modify the structure of the tests. Considering the pupils' perceptions may enrich not only the information about the standardized test, but it may also provide the tools to improve any deficiencies they might have.

Objectives and Research Question

The main objective of this research was to discover the perceptions the participants have on standardized English tests in the context of this study. The research question that guided this work was: what are the participants' perceptions of standardized English tests?

A Review of the Literature on Standardization in EFL

Assessing students' abilities and skills is a vital part of the education system. Evaluations are the most common method when assessing students' academic performance. As Çelik and Türkan (2014, p. 423) stated:

[...] evaluation involves the final product – what educators actually do with all of the information gathered from any assessment procedures they have employed; essentially, it is an objective judgment based on all of the evidence that has been assembled...t determines whether, and to what degree, each one of them (the students) as met the requirements and objectives of the course.

In other words, evaluations are the instruments with which teachers assess students' level of successfulness at the end of a determined course or subject. However, Kizlik (2012) views evaluation as a process, which gives information about factors such as "worthiness, appropriateness, goodness, validity, legality, etc" (p.2). He states that a reliable assessment must be made to obtain this information in order to allow teachers to know if students are learning appropriately. If the assessment is reliable, it might provide the level of knowledge the students have acquired.

As a means to assess students' knowledge of a particular subject, a teacher or school can choose from a variety of evaluation types. Gafoor (2013) mentions that the different types of evaluations "refer to different processes, products and the persons subject to evaluation" (p.2), and that, different data needs to be collected accordingly. Owing to this, it is important that schools decide the appropriate way of assessing students. In general, teachers can choose from two types of evaluations: formative and summative.

Formative and Summative Evaluations

The purpose of this type of evaluation is to point out students' weak areas in order for them to improve (Gafoor, 2013). In other words, "formative evaluation is diagnostic" (Udoukpong & Okon, 2012, p. 206). This type of assessment helps teachers to identify students' weaknesses as well as their strengths. If teachers are aware of this, they will be able to guide students throughout their learning process successfully. According to Gafoor (2013), some examples on how to assess students with this type of evaluation are weekly or monthly tests. These are common methods for teachers to become aware of their pupils' learning progress. In addition, Bognar and Bungić (2014) state that this kind of evaluation helps teaching to become better and richer. This is possible because as mentioned before, the purpose of this assessment is to determine students' weaknesses and help them to improve.

Summative evaluation is carried at the end of a course in order to determine if the objectives have been achieved properly (Gafoor, 2013). States et al. (2018) mention that summative evaluation “compares student knowledge or skills against standards or benchmarks” (p.1). In other words, the purpose of this evaluation is to discover if pupils have achieved the learning goals. Contrary to formative evaluation, which makes a diagnosis about students' learning, summative evaluation assesses the students' mastery of a determined subject or course.

There are diverse forms of summative evaluation. According to States et al. (2018), these forms include “midterm exams, final projects, papers, teacher-designed tests, standardized tests, and high-stakes tests” (p.1). These authors agree that standardized tests play a crucial part in summative evaluation, and that over the past 20 years, their role in schools has been increasing. Therefore, it is important to have an in-depth understanding of their impact and function.

Standardized Tests

When evaluating students, each school and/or teacher decides what kind of tests suits the context in which it will be administered. Standardized tests are one of the many options by means of which students may be evaluated, however, they have become a common way of evaluating. Benavides (2015) mentions that standardization became a widely accepted procedure in the beginning of the 70's in North America and Europe.

Standardized tests are administered and interpreted by using a uniform format (Bond, 1996). In other words, the purpose of these tests is to rate students' knowledge based on a standard. In addition to measuring students uniformly, the U.S Congress, Office of Technology Assessment (1992) explains that a standardized test (ST) “uses uniform procedures for administering and scoring. This ensures that scores obtained by different people are comparable to one another” (p.166). This means that standardization is a useful tool for comparing and

contrasting students' achievements in a determined subject or course. Furthermore, Kempf (2016) believes that STs rely on the assumption that comparing students with one another is needed so they can be ranked for diverse purposes.

As previously explained, the objective of standardized tests is to compare and interpret students' scores. However, this comparison can serve different purposes as well. Standardized tests can be divided into two categories, norm-referenced and criterion-referenced tests. Bond (1996) states that norm-referenced tests help to classify students. Based on these tests, students may be placed in different school programs, e.g., remedial or gifted programs. On the other hand, she defines criterion-referenced tests as those that report students' performance; whether they are doing well or not. In this study, criterion-referenced standardized tests is the category that is analyzed.

Design of Standardized Tests

The design of a standardized test has an important role to play in the assessment of students' gains. Standardization is generally used to measure students' knowledge to pass a course or to enroll in a new one. Therefore, it must be properly designed to fulfill its purpose. As Zucker (2003) mentions, "a well-designed standardized test offers a relatively affordable and efficient way of measuring the achievement of a large number of students." (p.4). The design of standardized tests has different parameters that have to be met for them to be successful when evaluating. These features are reliability, validity, and fairness. They are discussed in the following sections.

Reliability. This is an important element of standardized tests. Taylor (2013) explains that reliability "refers to whether students of roughly the same ability will consistently score similarly on the tests used" (p.6). In other words, reliability is related to the consistency of students' scores in a determined test. In addition, the U.S. Congress, Office of Technology Assessment (1992) mentions that reliability is

a function of the length of the test because the more items it has, the more reliable it is. Therefore, if a test is too short, it will be challenging to estimate the students' real knowledge.

Furthermore, the U.S. Congress, Office of Technology Assessment (1992) states that one way to conceptualize reliability is through the stability of the test. This means that if the test is administered on different days, students' scores will be consistent in all the attempts. For the U.S. Department of Labor Employment and Training Administration (1999) "a test that yields similar scores for a person who repeats the test is said to measure a characteristic reliably" (p.32). Reliability has a crucial role in the design of tests; it establishes that the test must give consistent results in order to be efficient.

Validity. The second important factor related to standardization is validity. The U.S. Congress, Office of Technology Assessment (1992) states that:

It is judged on a wide array of evidence and is directly related to the purposes of the test. Every test needs a clear specification of what it is supposed to be assessing. So, for example, for a test of reading proficiency, test designers first need to specify clearly what is meant by reading proficiency.
(p.177)

In other words, it is important that the test specify the purposes and competences that it covers. For the U.S. Department of Labor Employment and Training Administration, validity is the most important issue in selecting a test, as it "refers to what characteristic the test measures and how well the test measures that characteristic" (1999, p.35). They also emphasize the difference between validity and reliability; validity "will tell you how good a test is for a particular situation; reliability will tell you how trustworthy a score on that test will be." (p.36)

Additionally, The U.S. Congress, Office of Technology Assessment (1992) also mentions that validity is concerned with the fact that the test really measures what it claims to measure. Validity therefore is a key factor when designing a standardized test because the test needs to measure what has been specified.

Fairness. Fairness is related to validity, Kane (2010) states that “an assessment that is not valid in the sense that it tends to generate misleading conclusions or inappropriate decisions for some individuals or groups will also be unfair” (p.181). In other words, fairness means that rules and scoring should be clear and uniform in order for students to perform properly without these factors affecting them. For Camili (2006), fairness in testing refers to “perspectives on the ways that scores from tests or items are interpreted in the process of evaluating test takers for a selection or classification decision and it is closely related to test validity” (p. 225). A test is fair if it is clear and concise and if it is appropriate for all test takers because it does not mislead or affect their performance.

Fairness not only refers to the test itself; another aspect to consider is whether all the students covered the same content during the course, if it is not the case, then there will be unfairness. Lam (1995) states assessment is unfair if “students are not provided with equal opportunity to demonstrate what they know (e.g., some students were not adequately prepared to perform a type of assessment task) and thus the assessments are biased” (p.2). This is important since all the students will take the same test regardless of their preparation.

This situation has a major role to play. If students were not fairly prepared for the evaluation, it will affect their performance significantly. Lam (1995) adds that “these assessments are used to judge student capabilities and needs; and these distorted views of the students are used to make educational decisions that ultimately lead to limitations of educational opportunities for them” (p. 2)

Standardized EFL Tests

Standardized EFL tests belong to the category of criterion-referenced tests. The purpose of these tests is to identify whether students are achieving the goal of the course. The main objective of EFL courses is for students to reach a determined level of the language. According to Bond, criterion-referenced standardized tests are chosen when the objective is that:

Educators or policy makers wish to see how well students have learned the knowledge and skills which they are expected to have mastered. This information may be used as one piece of information to determine how well the student is learning the desired curriculum. (Bond, 1996, p. 2)

Standardized EFL tests fit this definition because the ultimate goal of the courses is that students master the language. The information these tests provide helps to determine if pupils have learned successfully or if they are having difficulties achieving the goals of the curriculum. Benavides (2015) observes that the results of standardized tests in languages are the basis on which inferences might be made about a person's level of knowledge.

Advantages of Standardized Tests

There are different reasons why standardization is often considered as the best option when evaluating. The main reason is they reduce class content. The National Council of Teachers of English (2014) observes that these evaluations “narrow the entire curriculum in many schools”, and that teachers “focus their instruction on the literacy skills measured on standardized tests” (p.2).

Additionally, the U.S. Congress, Office of Technology Assessment (1992) states that standardization permits to control the conditions in which the test is given and that, all students have the same instructions, materials and amount of time when

answering the test. This permits the reduction of “the effects of extraneous variables on a student’s score” (p.173). This is an important factor to consider; standardization allows students to have the same opportunity because all of them are answering the tests under the same circumstances.

Another important advantage of these tests as the U.S. Congress, Office of Technology Assessment (1992) explains is that they can be administered to large groups at the same time, and that it is not necessary for the examiner to guide the students because the instructions are clear and simple. This a good option for very large classes as, owing to the standard design, the test itself guides the students so there is no need for examiners to intervene. In addition, standardized tests are a “ready-made previously validated product that frees the teacher from having to spend hours creating a test” (Brown & Abeywickrama, 2010, p.104) In other words, these tests are practical because they are already available for teachers.

Finally, the U.S. Congress, Office of Technology Assessment (1992) also states that this procedure ensures that “an individual receives a score that reflects his or her level of understanding and not the particular views or attitudes of persons administering or scoring the test” (p.175). This is a key factor because standardization shows students’ scores based on their knowledge and not on someone else’s perceptions.

Disadvantages of standardized tests

Standardization has a variety of advantages, but some authors point their deficiency in certain aspects. The National Council of Teachers of English (NCTE) (2014) argues that standard tests limit students’ learning because instead of focusing on different qualities, they focus only on cognitive dimensions. The NCTE continues by explaining that teachers focus simply on drills for the tests instead of helping students develop non-cognitive abilities. This is a major problem with standardization: teachers prepare students for the tests and for the specific topics

that compose it, forgetting about other activities and abilities that could help students as well. This factor is known as the washback effect. Buck (1998) defines it as a “natural tendency for both teachers and students to tailor their classroom activities to the demands of the test, especially when the test is very important to the future of the students” (p. 17).

When standardization is the method by which students' knowledge is measured, the content of the lessons is aimed at preparing the students to pass the examinations. Taylor (2013) states another disadvantage of standardization. He mentions that STs encourage superficial learning; not the mastering of a determined subject or topic and that, what students learn is often not retained after the exam. As stated previously, students focus only on the content of the tests and sometimes they simply memorize that content, instead of actually learning. This is an important factor to consider when choosing this type of test; teachers must be careful that students learn the content so that they know it even after the test has taken place.

Another disadvantage to consider, according to Brown and Abeywickrama (2010) is that in the case of multiple-choice tests, test takers firmly believe that these questions are valid and reliable. They mention that these instruments appear to have authority, but they might not necessarily have an “appropriate construct of validation” (p.105). Not all standardized tests are valid and reliable; they must serve these purposes in order to be efficient; otherwise, it would be a useless measure students' abilities and knowledge with them.

Methodology

For this research, a questionnaire was used as method of collecting information. It contained open-ended questions, which give an insight of the students' viewpoints. However, it had close-ended questions as well which provided numerical data that could be used to support students' perceptions. Thus, even though the main

objective was to discover what students think of standardization, percentages were also used in order to support the results and the participants' answers.

This type of data collection had some advantages that were useful for this study. It helped to collect a large amount of information in less time than interviews. In addition, processing the data was faster and easier. This was deemed the best option for this research because the number of participants was large, and it was necessary to have them answering the same questions at the same time.

For this questionnaire, open-ended questions were used, these “allow[ed] the respondent to express an opinion without being influenced by the researcher” (Foddy, 1993, p. 27). Other advantages of using this type of questions was that they helped to discover spontaneous answers given by the participants and it helped to prevent the researcher from suggesting the responses (Reja et al., 2003).

Context

The research of this case study was conducted at the language department of a university in Mexico. It is a public school and it is ranked amongst the best universities in the country. This school uses a program called MEIF (the acronym in Spanish for Comprehensive Flexible Educational Model). In theory, it allows students to choose their schedules, their subjects, their teachers, and the number of terms in which they want to finish their degree. The goal of the English degree, which was the object of study in this research, is to help students learn English as a foreign language.

There are six levels for the English courses in this institution, which are the following: beginners, elementary, pre-intermediate, intermediate, upper-intermediate, and advanced English. In order to measure students' level of English, this school evaluates students by means of standardized tests. Students are evaluated three times during the semester; the first two tests are administered to all students the

same day at the same time. As for the final exam, the students take their test on different days and have to follow a schedule. Testing takes longer because there is a writing part. They have to take an oral evaluation (in pairs) as well. These standardized tests are thought to measure the four English skills of students (listening, writing, speaking, and reading).

The tests account for 90% of the total score; the other 10% is accounted for by a portfolio, which is determined by each teacher. The standardized English tests have an important role in this BA, as they account for the larger part of the students' score. Therefore, if students fail the three tests, they must repeat the course and, if the course is failed twice, they will be dismissed from the university and shall not be readmitted to the same program.

Participants

The participants for this research were advanced students of the English BA at a university in Mexico. Advanced students were chosen because they have already experienced five previous courses and have answered the standardized test on several occasions. These students were familiar with this evaluation process because they had almost finished all the English levels, therefore their views provided useful information regarding standardization.

Two advanced English groups were the participants for this research. The two groups have a total of 52 students and their ages are between 21 and 24 years old. As stated before, they are in the last level of the English courses, and this is the last term in which they were evaluated using the standardized test. It is important to consider their perceptions because they are aware of the structure and purpose of standardization.

Results and Discussion

This section is about students' perceptions on a) the design of the standardized tests, b) validity, c) fairness, and d) reliability. Some of the issues herein discussed include instructions, structure, the washback effect, memorization, class content, scores consistency, and factors that influence reliability.

Design

Design has an essential role in tests, as Camilli (2006) mentions, "Tests designed properly and used fairly can facilitate positive educational, social, and economic goals" (p.225). In addition, the Educational Testing Service (2014, p.29) states that "developers should know what the test is intended to measure, the characteristics of the intended test takers, and how the test is intended to be used."

Instructions.

An important part of the design of a test are instructions because they guide students throughout the evaluation. As Sowell (2017) states, instructions "have a direct effect on learning; a lesson or activity becomes chaotic and fails when students do not understand what they are supposed to do" (p.10). In this research, students were asked about their thoughts regarding instructions. 63% of the students agreed that questions are short, clear, precise, and easy to understand. As the percentage shows, for the majority of students, instructions are not a problem. For example, Gerardo shared:

- *They're not long nor difficult, they are very clear to understand.*

In this particular case, instructions do not seem to be a problem in the design of tests. Students perceive them as clear and easy to understand. Most of them, agree there is an appropriate structure in the instructions.

Appropriate Design.

Regarding the design of tests, when students were asked if tests are well elaborated (this means they are precise, without mistakes, ambiguities, or confusing parts) most of them responded they do not believe tests are well designed. Some of the reasons students stated as a lack of a successful design, were the ambiguity of questions, the errors they sometimes have, and the fact they only admit one right answer, even though there are different answers that have the same meaning and are all correct. For example, Magdalena pointed out:

- *I don't consider tests well-elaborated because they have some mistakes and questions are ambiguous.*

Students' answers suggest that most of them do not consider tests to be properly designed.

Structure.

One important point regarding test design is the topics they decide to include and the abilities they evaluate. In these standardized tests, only the final exam evaluates the four language skills, while the other two just focus on grammar and vocabulary. 69% of students stated they do not agree with this, they would like these abilities to be evaluated throughout the three tests because they consider them important. The students that do not agree on this choice of design stated they consider these 4 skills fundamental while learning a language. They believe the development of these abilities helps them practice the language, be more fluent and proficient and learn from their mistakes. For example, Andy mentioned that:

- *The four language skills should be constantly evaluated because students need to practice in order to correct their mistakes.*

On the basis of these testimonies, it can be safely assumed that students would like the four abilities to be constantly evaluated.

Validity

This section explores student's perceptions regarding the test's validity. This is a key factor in standardized tests, as mentioned by the U.S. Department of Labor Employment and Training Administration (1999): "validity is the most important issue in selecting a test. It refers to what characteristic the test measures and how well the test measures that characteristic" (p.25). In other words, validity ensures that the specific test purpose is being measured appropriately.

Valid Evaluations.

One of the questions the participants were asked was if they believed the tests really evaluated what they had seen in class. This is important because the class content is what is evaluated in the first two tests. 44% of students stated they partially agreed with this. One of the reasons they gave was that the tests evaluate the lessons seen, however sometimes there is not enough time during the course to see all the content. Camila shared:

- *Tests partially reflect what we learned in class, each teacher teaches differently and at a different pace. Sometimes not all the topics are covered and they come on the test but we didn't see them.*

Tests evaluate what they said they would, however not all the students are exposed to the same content.

Washback Effect.

Tests focus on determined topics and themes; consequently, it is common that lessons are developed to help students become more familiar with the test

structure and that content. However, students consider that this is not necessarily the best idea. One question was whether students believed their tests' scores reflected their ability to communicate in English. 65% of the students answered no. For example, Sandra claimed that:

- *Classes being limited to answering the student's book is not enough to learn because tests evaluate the book content, all the units from the book come on the test, in the classes we only solve the book's activities. There are 12 units on the book and that's all you need to learn so you can pass, in the end it is not really meaningful.*

This belief could be related to the washback effect. Buck (1998, p. 17) defines it as “a natural tendency for both teachers and students to tailor their classroom activities to the demands of the test, especially when the test is very important to the future of the students.” In this context, the standardized test is fundamental for students. They need to pass the tests in order to move to the next English level and to be able to continue studying their BA. However, as Sandra mentioned, the fact that all lessons revolve around the test content does not provide meaningful learning because students only need to focus on the book, although sometimes they want to learn more than that.

In addition, some students state that answering tests on the content of a book does not really show their English proficiency. Andy commented:

- *Both the tests and the classes focus on the book content, however, sometimes there is not enough time to see all the units and sometimes the content does not match what we saw during our classes, so it does not really show how good we are in English.*

According to students' perceptions, lessons are merely seen as preparation for a test owing to its importance in an academic context. This is why lessons appear to be only about the book content. This then seems to echo Kempf's (2016, p.19) view that "it is challenging to think of other activities that play an equally central role in daily life in schools; and remember this is not teaching content but rather test preparation."

In addition, Kempf (2016) mentions that teachers are asked to do this test preparation even if the agenda or curriculum is already packed. Consequently, as the evaluations are based on book content, lessons also need to be based on that same content in order to prepare students to be successful in their tests. However, as Andy mentioned:

- *Sometimes there is not enough time to see all the units.*

Tests are based on the class, book content, however, the purpose they have is to measure student's English proficiency, and many students agreed that measuring their knowledge on the book content was not really measuring their English level.

Memorization.

Some participants appeared to agree that learning is not meaningful for them because they only learn to pass the test. For example, Estela explained that:

- *Learning is not meaningful because students only need to focus on book vocabulary and grammar, so it is not relevant for students because it does not go beyond that; you only study the content to have a good score.*

Students agree that learning becomes meaningless to them because their only purpose is to pass the test, not necessarily learn the content. These is broadly in line

with Taylor's (2013) findings. In his research context, he found that 90% of the students agreed that they only focus on test content rather than mastering the subject or enjoying learning, and 75% of these students reported they "do not retain information learnt for a long time after the exam" (Taylor, 2013, p.23).

In this research project, participants agreed that their learning is not permanent. As Hugo mentioned:

- *Learning is superficial, you only do it for the test and then it isn't important to you anymore because you just used that knowledge for passing the quiz.*

This also appears to be related to Taylor's (2013) research. He mentions that standardization encourages superficial learning not the mastering of a determined subject or topic and that, what students learn is often not retained after the exam.

In other words, if learning is not relevant for students after the test, they will simply forget about it, which could mean that it was meaningless for them. As stated previously, students focus only on the content of the tests and sometimes they simply memorize that content instead of actually learning. It is therefore common that their knowledge does not remain in their memory because they only learn for the test.

Fairness

Fairness is an important aspect regarding standardization. All the students take the same evaluation; however, there are different factors that can lead to unfairness.

For the Educational Testing Service (2014) "all test takers should be treated comparably in the test administration and scoring process" (p. 20). In this section, I discuss whether students consider standardized test fair.

Class Content.

Participants were asked whether they believed standardized tests were fair in terms of whether they evaluate what teachers taught in class, 52% disagreed. For example, Amanda stated:

- *Yes, it evaluates the content seen in the book and you have the book to study, but sometimes you don't see that content in classes.*

In other words, tests must evaluate class content; however, the problem is that sometimes this content is not seen in classes. This leads to unfairness because not all the students are under the same conditions when taking the evaluation. It is also important to consider that students must pass the tests in order to continue in university. One of the students, Hugo, pointed out:

- *Each teacher teaches differently and at a different rhythm, sometimes not all topics are covered, and they come on the test, sometimes you feel on your own.*

As the participants mentioned, there is inconsistency in the class content amongst the different groups. As the Educational Testing Service (2014) states: "the best way to approach the ideal of fairness to all test takers is to make the influence of construct-irrelevant score variance as small as possible" (p. 19). It is important to consider all the variables that can affect students' performance. Marie explained:

- *Each teacher gives you different content, and the head teacher does not create the tests and only he knows what we saw in class.*

The main issue seems to be that, although the test might be considered fair, in terms of administration and course content, not all the students are exposed to the same content in class.

Reliability

Reliability is a key concept when designing tests. According to Hale & Astolfi (2014) reliability is “an indicator of consistency, i.e., an indicator of how stable a test score or data is across applications or time. A measure should produce similar or the same results consistently if it measures the same thing.” (p. 45). In other words, it is about consistency in the students' scores. Additionally, De Bruin (2010) states that “a test is seen as being reliable when it can be used by a number of different researchers under stable conditions, with consistent results and the results not varying.” (p. 54).

Scores Consistency.

As mentioned previously, reliability reflects consistency. The Educational Testing Service (2014) explains that reliability “refers to the extent to which scores (or other reported results) on a test are consistent across— and can be generalized.” (p. 26). Participants were asked about their scores. They chose between five different options. These options were divided in different score ranges, on a scale from 5 to 10. The results showed that 48% of the students define their score as “good”. The participants stated they (48%) mostly score between 7.5 and 8.4. The category that follows with 34% is “low” which means they get scores between 6 and 7.4.

The results from this question show that students seem to be getting the same range of scores when taking the evaluation. Participants agreed that when they take the standardized test, their score is the same. This score is neither the best nor the highest, but students are not failing; therefore, the test appears to be serving its purpose. Melisa mentioned:

- *My score is always the same, I get an 8, it's not the best but I'm passing and that's what matters.*

Nevertheless, some students mentioned they wish the results were better. Hugo explained:

- *My score is good, I usually get an 8 but I know it could be better if we checked all the content or if the structure of the test was different.*

Factors Influencing Reliability.

Reliability depends on stability; nevertheless, there are factors that can interfere with it. The U.S. Department of Labor Employment and Training Administration (1999) mention that “test performance can be influenced by a person's psychological or physical state at the time of testing. For example, differing levels of anxiety, fatigue, or motivation may affect the applicant's test results” (p. 22). One of the questions I asked students was how they felt when taking the evaluation. They had a list of 12 different feelings: positive and negative. They could choose more than one option if they wanted. The highest percentages were negative feelings such as nervousness, concern, stress, and pressure. 63.5% of the students agreed they feel nervous when taking their evaluations. For example, Amanda pointed out:

- *I feel nervous because I know sometimes there is content the teacher didn't teach.*

It is important to notice this factor because it might affect students' performance since most of them have negative feelings when solving their test. Hale and Astolfi (2014) state that “poorly motivated, fatigued, uncomfortable examinees will perform below their capability [...] everything possible should be done to motivate examinees” (p. 49).

Conclusion

The purpose of this research was to find out students' perceptions regarding standardization in English tests. Conclusions are based on qualitative and quantitative information collected from participants. There were three concepts that were considered: design, validity, reliability, and fairness. Regarding design, overall, students agreed that instructions are clear; however, they pointed some ambiguities in questions. A key factor is that participants do not agree with the fact that tests only evaluate the four language skills in the final test. In conclusion, participants hope for a different test structure that could assess these skills in the three exams. Thus, their classes can stop focusing on learning grammar and vocabulary.

Another important factor was students' perceptions regarding test validity. Participants partially agreed with tests being valid in evaluating content they actually covered in classes. The constant in students' answers was that it depends on the teacher because not all of them teach or cover all the content. Two important factors emerged from students' perceptions: the wash-back effect and memorization. Such factors have been considered by different authors as the main disadvantages of standardization. Classes revolve around the book content because tests are based on it. In addition, students do not consider their learning meaningful because they simply memorize book content in order to pass. Thus, validity seems to be arguable; tests evaluate what they should (English proficiency based on class content); however, not all the students are exposed to the same content.

Validity is intrinsically tied to fairness. The participants stated they do not feel tests are fair when evaluating class content due to different teaching styles. These standardized tests do not appear to be fair because not all students are under the same conditions. This factor should change because students depend on these tests to continue in the BA. Regarding reliability, tests appear to be serving its purpose. Students agreed their score seems to be consistent every time they take the test. Nevertheless, these scores are not the highest; students seem to have a low level.

One of the reasons could be the negative feelings they have before the evaluation. Most of them feel nervous, stressed, or worried; this could be affecting their actual performance level.

Standardization plays an important role in these students' academic life. The tests account for 90% of the total grade. Therefore, more attention should be given to the design, validity, and fairness. It seems that these tests need some adjustments. It is true the tests are evaluating the book topics; however, there should be concern about the fact that not all teachers are covering all the contents. In addition, according to students' perception, there should be a change in the structure. Students wish there were more attention given to the four language skills and not only to grammar and vocabulary. They want their learning to go beyond a book and be meaningful.

To comprehend more the role and effectiveness of standardization, there needs to be more research carried out in the other English levels. It would be useful to find out how students feel as they are promoted to higher levels and semesters. Another important factor that can be investigated more deeply is class content. It is essential to discover why some teachers do not complete contents or why they skip them, as this is affecting students' performance on tests. Finally, there is a need for more research about students' feelings regarding evaluations and to what extent they truly affect their scores.

The purpose of this research was to contribute to the study of the effectiveness of standardization based on students' perceptions. In paper, standardized tests might appear as the best solution for big groups. Nevertheless, it is essential that they are well-structured, valid, and most important, that every group be exposed to the same conditions so that it is fair to take the same evaluation.

Limitations

One of the limitations of this study was that the sample of participants was limited to two advanced English courses. Perhaps a more sample might have yielded more conclusive patters in the results. Moreover, the sample could have included students from the lower courses to find out if perceptions varied across the different levels. As a result, results cannot be generalized to other contexts and populations.

Another important limitation was that we administered a survey in which there was an open-ended questions section. Perhaps, using additional data collection tools would have given us another perspective on the problem we set out to investigate.

References

- Benavides B., Jorge E. (2015). Las Pruebas Estandarizadas como Forma de Medición del Nivel de Inglés en la Educación Colombiana. In J. Bastidas, & G. Muñoz (Eds.), *Fundamentos para el desarrollo profesional de los profesores de inglés* (pp. 29-56). Editorial Universitaria – Universidad de Nariño.
- Bognar, B., & Bungić, M. (2014). Evaluation in Higher Education. *Život i škola*, 1(31), 139 –159.
- Bond, L. (1996). Norm- and Criterion-Referenced Testing. *Practical Assessment, Research, and Evaluation*, 5(2), 1-5.
- Brown, H. D., and Abeywickrama, P. (2010). *Language Assessment: Principles and classroom practices*. Pearson Education.
- Buck, G. (1998). Testing Listening Comprehension in Japanese University Entrance Examinations. *JALT Journal*, 10(1), 15-42.
- Camilli, G. (2006). Test fairness. In R. Brennan (Ed.), *Educational Measurement* (pp. 221-256). American Council on Education.
- Çelik, S., & Türkan, S. (2014). Assessment and evaluation in EFL: Classroom methods, assessment tools, and the instructional implications of high-stakes exams. In S. Çelik (Ed.), *Approaches and Principles in English as a Foreign Language (EFL) Education* (pp. 419-428). Egiten.
- De Bruin, I. (2010). Exploring how objects used in a Picture Vocabulary Test Influence Validity. Dissertation (MEd)--University of Pretoria.

Department for Education in the United Kingdom. (2011). The Framework for the National Curriculum: A Report by the Expert Panel for the National Curriculum Review. Department for education.

Educational Testing Service. (2014). ETS Standards for Quality and Fairness. Educational Testing Service.

Foddy, W. (1993). Constructing Questions for Interviews and Questionnaires: Theory and Practice in Social Research. Cambridge University Press.

Gafoor, K. (2013, 4 9). Evaluation in Educational Practice Types and Phases. Retrieved from https://www.researchgate.net/publication/272237015_Types_and_Phases_of_Evaluation_in_Educational_Practice

Graham, C., & Neu, D. (2004). Standardized testing and the construction of governable persons. *Journal of Curriculum Studies*, 295-319.

Hale, C., & Astolfi, D. (2014). *Measuring Learning and Performance: A Primer*. Saint Leo University.

Kane, M. (2010). Validity and fairness. *Language Testing*, 27(2), 177–182.

Kempf, A. (2016). *The Pedagogy of Standardized Testing*. Palgrave Macmillan US.

Kizlik.B. (2012). *Measurement, Assessment and Evaluation in Education*. Robert Kizlik and Associates Boca Raton, Florida. Retrieved From <Http/Www.Adprima.Com> Measurement

Lam, T. (1995). *Fairness in Performance Assessment*. Greensboro, NC: ERIC Clearinghouse on Counseling and Student Services. ERIC Document Reproduction Service No. ED39198.

- National Council of Teachers of English. (2014). How Standardized Tests Shape—and Limit—Student Learning. *The Council Chronicle*, 24(2), 1-3.
- Reja, U., Manfreda, K. L., Hlebec, V., & Vehovar, V. (2003). Open-ended vs. Close-ended Questions in Web Questionnaires. *Advances in methodology and statistics*, 19(1), 159-177.
- Rocha, A. (2014). Students' perceptions regarding standardized English tests: The case of beginner EFL students (unpublished bachelor's dissertation). Universidad Veracruzana.
- Sowell, J. (2017). Good Instruction-Giving in the Second-Language Classroom. *English Teaching Forum*, 55(3), 10-19.
- States, J., Detrich, R. & Keyworth, R. (2018). Overview of Summative Assessment. Retrieved from <https://www.winginstitute.org/assessment-summative>.
- Taylor, M. (2013). Student perceptions of standardised testing: Survey exploring the attitudes of university undergraduate students towards the traditional methods of student evaluation in formal education. Lancaster University.
- U.S. Congress, Office of Technology Assessment. (1992). *Testing in American Schools: Asking the Right Questions*. U.S. Government Printing Office.
- U.S. Department of Labor Employment and Training Administration. (1999). *Testing and assessment: an employer's guide to good practices*. U.S. Department of Labor, Employment and Training Administration, Office of Policy and Research.
- Udoukpong, B., & Okon, C. (2012). Perception of Formative Evaluation Practices and Students' Academic Performance In Junior Secondary Certificate

Examination in Social Studies. *International Journal of Business and Social Science*, 3(15), 204-212.

Zucker, S. (2003). *Fundamentals of Standardized Testing*. Pearson Education.