

PAPER • OPEN ACCESS

Application of the Wilcoxon Test to correlate the results of the Saber 11 and Saber T&T Test

To cite this article: L K Ahumada and I D Sanchez 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **519** 012034

View the [article online](#) for updates and enhancements.

Application of the Wilcoxon Test to correlate the results of the Saber 11 and Saber T&T Test

L K Ahumada¹ and I D Sanchez¹

¹Antonio de Arevalo Technological Foundation TECNAR

Luz.ahumada@tecnar.edu.co

Abstract. In Colombia there is a model of monitoring the Quality of Education in which different tests are applied throughout the academic lifecycle of a student, in basic education, secondary education, until finally reaching higher education. Among these tests are: the Saber Test 11 and the Saber T & T Test, among others. The first, applies to students of the last grade of secondary education and the second to students of the last semester of the Technical and Technological program. It is possible to think that a person who obtained good results in the test saber 11, after preparing in a Higher Education program and presenting the corresponding test, should present a notable improvement in their results, if compared with the first exam. The foregoing is not necessarily true, in order to make this statement it is essential to carry out a study that supports this theory. The present research makes use of Test Wilcoxon to compare the means of 2 distributions, for this case particularly: the Saber 11 and Saber T & T tests of the same student. From the Saber 11 test, the generic Mathematical and language competences are chosen, and for the Saber T & T tests, quantitative reasoning and critical reading were selected. The selected population are the students of the Technology in Systems program who took the test in 2012, belonging to the Faculty of Design and Engineering of the Technological Foundation Antonio de Arevalo - TECNAR, they are students, product of the first accreditation of the program.

1. Introduction

In Colombia, education is defined as a process of permanent, personal, cultural and social formation that is based on an integral conception of the human person, their dignity, their rights and their duties. [1]

The Colombian educational system is made up of: initial education, pre-school education, basic education (in primary, five grades; and in secondary, four grades), secondary education (two degrees and culminates with a bachelor's degree) and higher education [2]. The latter was selected to carry out the investigation.

Quality, in the context of education, and especially applied to institutions of higher education, refers to an attribute of the public service of education in general and, in particular, to the way in which that service is provided, according to the type of institution concerned [3]. Quality, well understood, entails the continuous effort that the state and its institutions make to fulfill with responsibility the specific demands of each one of its functions [4].

One of the mission objectives of the Ministry of National Education-MEN is to improve the quality of education regardless of what level we are talking about.



In order to measure quality, the MEN devised an instrument called Saber Tests, which is applied according to the degree of training in the following way:

Saber Test. Applies to 5th and 9th grade students. It focuses on areas related to Language, Mathematics, Natural Sciences and Citizen Competencies.

Saber Tests 11. Applies to students who finish grade 11 and focuses on areas such as language, math, physics, chemistry, biology, geography, history, philosophy, foreign language

Saber Tests T & T. It is applied to students who finish technical and professional programs in Higher Education Institutions. The learning and the competences of the academic program that is being studied are evaluated. The present investigation focuses on these last two tests, it should be mentioned that the presentation of these exams is obligatory and sometimes they are an essential requirement to obtain the degree to which they aspire.

2. Institutions of Higher Education in Colombia

The Institutions of Higher Education (IES) are entities that count, according to the legal norms, with official recognition as providers of the public service of higher education in the Colombian territory [5].

The academic character is the main feature that since the constitution (creation) of a higher education institution defines and gives identity to the competence (field of action) that in academics allows you to offer and develop higher education programs, in a another academic modality [6]

- Higher education in Colombia is divided according to its academic character into 4 levels:
- Professional Technical Institutions: These institutions that are characterized by offering short-term technical programs.
- Universities: those that are able to offer all academic programs including masters and doctorates.
- University Institutes: Are institutions that offer university degrees, graduates get titles as "professionals"
- Finally, the technological institutions: These are institutions that offer technological programs at the professional level in a medium term.

3. Characterization of the Program

Precisely, in this last category is where the Antonio de Arévalo Technological Foundation is located. It was created at the beginning of the 80s as a technological foundation in the city of Cartagena. [7]

Currently, TECNAR has three faculties: Faculty of Economics, Faculty of Social Sciences and Faculty of Design and Engineering. This latter has 14 academic programs, among which is the Systems Technology Program. Accredited on three successive occasions as a High Quality program, fulfilling the mission objectives established by the Ministry of National Education.

The Technology-in-Systems Program has been characterized by training students in the areas of software knowledge. in order that graduates can perform with a technological and humanistic approach in functions of: Development of computer programs, deploying skills and abilities for the use of technological means in an open and permanent manner also involves other areas of knowledge such as: exact sciences, communication skills among others.

To date, the program has approximately 340 graduates. The communication with them is given through the division of Social Projection, which is a specialized area that tends to maintain close and permanent relations with them.

Currently, the institution does not have a mechanism to measure the knowledge that graduates acquired during their permanence at the institution. Therefore, the academic units can not reflect or assess themselves about the relevance of the microcurriculum with strong bases. Due to the above, it is necessary to analyze the performance of the graduates of the Technology-in-Systems program. it is then suggested to make a comparison of the results related to the generic competence of the Saber Test 11 (The students in what state they enter the institution to study? academically speaking), Saber T & T (The students in what state they finish their studies in the institution? academically speaking)

4. Methodology

In order to obtain answers to the previous interrogations, we chose as a unit of analysis those high school graduates from the different schools of the city of Cartagena - Colombia, who later attended and later graduated from the Technology in Systems program. The databases used were downloaded from the Colombian Institute for Education Evaluation and the Institutional Information System portal. Subsequently, mathematical and statistical models were used to correlate variables and give a light in their behavior, with the Wilcoxon test being the best option to achieve it.

It must be taken into account that the Saber 11 test differs in the year of presentation and therefore in the structure and evaluation of it. To remedy that it is necessary to make a transformation in the scores of the Saber T & T Tests because the Score Range is 0.200 and you want to make the comparison with the Saber Tests 11 whose Range is 0.100. For this circumstance in the present investigation, the division of the test score will be done by T&T by 2 and thus parity will be made between the results of the 2 tests.

The Wilcoxon test is the tool used to perform the correlation, it is a statistical test that contrasts two samples of the qualitative variable relationship. [8] For the present study (Saber Tests) and another quantitative variable (results obtained by the students) and which makes a comparison contrast of the central tendency, the median.

Testing is used when there are very small sample spaces. Wilcoxon contrast can be used when a random sample of linked pairs is available [9]. In the present study they are related because the same students are evaluated in two different moments, a first moment with the Saber Test 11 and a second moment with the Saber Test T & T.

With this test, we will compare the differences of the averages of the Groups. Between the average of the Results of the Saber Test T & T, and Saber Test 11, and answer the Hypothesis if the differences between this median is zero, this is Null Hypothesis (H_0 : Difference between averages = 0) or if the differences between the averages is different from zero, this is the Alternative Hypothesis (H_a : Difference between averages $\neq 0$).

If the statistical result yields a result of **H_0** then it would be concluded that the results of the Saber T & T Tests and the Saber Tests 11 do not have a significant change after the process of Formation of the students in the Technology Systems Program. Otherwise. If the statistical result yields a result of **H_a** then it would be concluded that the results of the Saber T&T Tests and the Saber 11 Tests have a significant change after the process of training of the students in the Program. In the latter case, it would be defined if the change were positive or negative depending on which of the two averages is greater.

The sample used corresponds to 30, who made their Saber T&T Test in 2012. The Wilcoxon test is applied for data greater than 25 datas, where $n > 25$, and is calculated as follows:

- 1) The differences in each element of the sample are calculated for the two variables to be studied (Saber T&T, and Saber 11).
 - 2) Subsequently the differences are ordered by discarding the signs (+ or -).
 - 3) Then an order range is assigned to them. (The values that are repeated are not given different order ranges, we assign an average value).
 - 4) The ranges are added according to the signs that have the differences and the estimators are obtained:
 - a. $W(+)$: Sum of Ranges corresponding to positive differences.
 - b. $W(-)$: Sum of Ranges corresponding to negative differences.
 - 5) The Contrast Statistician is defined, which is used to make the acceptance decisions of the H_0 or **H_a** , that is, the minimum value between the two sums of ranges: $W = \min [W(+), W(-)]$
- Steps to follow when the sample is greater than 25 data, $n > 25$: Follow the same steps as before to point five.

6) As the statistician fits a normal distribution, the W statistic is transformed into a new value designated as Z with the formula (1):

$$Z = \frac{W - \frac{n(n+1)}{4}}{\frac{\sqrt{n(n+1)(2n+1)}}{2a}} = N(0,1) \quad (1)$$

8. This Z Value is compared to a Normal value $Z\alpha$, which we know for a 95% confidence level $Z\alpha = 1,96$. If the experimental value $|Z| < Z\alpha$ would accept the **H₀** and if experimental value $|Z| > Z\alpha$ then it reject the **H₀** and the **H_a** would be accepted.

5. Results

In Table 1 where there are related calculations with the Wilcoxon Test. Two-tailed test: Quantitative Reasoning of the Saber test T&T and Mathematics competence in Saper 11, it is observed that the result of the experimental value transformed "Z" is greater than the critical value of the Wilcoxon Table. Therefore, there would be a risk if the null hypothesis is rejected. For this reason, we can conclude that, since the differences between the medians are null, no statistical evidence has been found in the set that indicates that between the results of the Critical Reading tests and the Reading Comprehension tests there is no Change significant after the process of training students in the program for the academic period under study. Although it is observed that the transformed experimental value Z is lower than the critical value of the Wilcoxon Table, therefore, there is a high acceptance rate of the alternative hypothesis, bearing in mind that the change is highly significant for this case.

5.1 Quantitative Reasoning

This table shows that Quantitative Reasoning (QR) has a significant change of "positive" character in the training process that is to say that students after having lived their formative process reached to overcome with respect to the first test performed (Saber 11) , it is observed that when evaluating the student in the Saber T & T Tests compared to the initial ones, in the evaluated competence that for this case was QR Quantitative Reasoning, 66.6% of the students obtained a considerable improvement. Improvement in their performance in Saber tests T & T.

$$Z_{QR} > W_0 QR$$

5.2 Critical Reading

To the right of Table 2 the analytical summary of the results of the method for generic competence is observed. Critical reading, although there is no statistical evidence to reject the **H₀** null hypothesis, this does not indicate that there are negative changes in the results. It should be noted that 60% of the population obtained notable improvements in their performance despite not having achieved the desired results.

$$Z_{CR} < W_0 CR$$

Table 1. Calculations with Wilcoxon Test Two-tailed test - Saber T&T Quantitative Reasoning and Saber 11 in Mathematics

DQR	DCR	IDIFI RC	IDIFI LC	QRR	CRR	RRQR	RRCR	RSQR	RSRC	W+ RC	W- RC	W+ LC	W- LC
-13.66	0.65	13.66	0.65	26	1	26	1	-26	1	-26			1
-1.05	-1.17	1.05	1.17	3	2	3	2	-3	-2	-3		-2	
4.17	1.51	4.17	1.51	15	3	15	3	15	3		15		3
3.53	2.23	3.53	2.23	13	4	13	4	13	4		13		4
-3.64	-2.78	3.64	2.78	14	5	14	5	-14	-5	-14		-5	
1.67	-2.99	1.67	2.99	4	6	4	6	4	-6		4	-6	
2.17	-3.11	2.17	3.11	7	7	7	7	7	-7		7	-7	
-7.54	3.41	7.54	3.41	18	8	18	8	-18	8	-18			8
-8.73	-3.57	8.73	3.57	20	9	20	9	-20	-9	-20		-9	
-11.12	4.22	11.12	4.22	23	10	23	10	-23	10	-23			10
-10.5	-4.5	10.5	4.5	21	11	21	11	-21	-11	-21		-11	
-14.7	-4.9	14.7	4.9	27	12	27	12	-27	-12	-27		-12	
-4.55	-6.15	4.55	6.15	16	13	16	13	-16	-13	-16		-13	
3	6.95	3	6.95	11	14	11.5	14	11	14		11		14
-2.5	-7.5	2.5	7.5	8	15	8.5	15.5	-8	-15.5	-8		-15.5	
-3	-7.5	3	7.5	12	16	11.5	15.5	-12	-15.5	-12		-15.5	
-10.77	-7.58	10.77	7.58	22	17	22	17	-22	-17	-22		-17	
-7.85	-7.86	7.85	7.86	19	18	19	18	-19	-18	-19		-18	
2.5	8.5	2.5	8.5	9	19	8.5	19	9	19		9		19
-18.12	-9.24	18.12	9.24	29	20	29	20	-29	-20	-29		-20	
-18.23	-10.59	18.23	10.59	30	21	30	21	-30	-21	-30		-21	
0.19	10.73	0.19	10.73	1	22	1.5	22	1	22		1		22
-2.97	-10.9	2.97	10.9	10	23	10	23	-10	-23	-10		-23	
-12.62	11.75	12.62	11.75	25	24	25	24	-25	24	-25			24
0.19	12.3	0.19	12.3	2	25	1.5	25	2	25		2		25
-17.44	-12.4	17.44	12.4	28	26	28	26	-28	-26	-28		-26	
5.64	13.14	5.64	13.14	17	27	17	27	17	27		17		27
-12.5	-14.5	12.5	14.5	24	28	24	28	-24	-28	-24		-28	
2.05	17.77	2.05	17.77	6	29	6	29	6	29		6		29
-1.99	-19.72	1.99	19.72	5	30	5	30	-5	-30	-5		-30	
										-380	85	-279	186

Where: DQR = Difference in quantitative reasoning, DCR = Difference Critical Reading, QRR = Quantitative reasoning range, CRR = Critical reading range, RRQR = Range real of quantitative reasoning, RRCR = Real range of critical reading, RSQR = Range with sign quantitative reasoning, RSRC = Range with sign reading critical.

Table 2 show the analytical summary of the results of the method:

Table 2 Analytical summary of the results of the method

QR				CR			
W RC	N	Z RC=	Wo RC	W LC	N	Z RC=	Wo LC
85	30	3,03	1,64	186	30	0,956	1,64

Table 3 Analytical summary of the results of the method

QR			CR		
T	N	Z	T	N	Z
85	30	-3,03	186	30	-0,96

The results shown in this research are susceptible to improvement. It is suggested to make curricular changes and trainings that tend to strengthen each of the generic competences in the level of teaching and the student body. For which it is important to take into account the following topics:

In Quantitative Reasoning: Interpretation of Tables and Charts in which it involves numerical information. Justification of procedures and strategies in problem solving. Basic topics such as Percentages and probability of elementary events and events.

In Critical Reading: Reading of texts that generate reflection, understanding and criticism. Find the general meaning of a text and make inferences from it.

It is important to take a diagnostic entrance exam, so that students know the status of generic competences.

Finally, carry out an institutional program that starts from the first semester to strengthen these Competitions

It is important to perform a diagnostic entrance exam to the students to know the status in the generic competences. Carry out an Institutional program that starts from the first semester to strengthen these competences.

6. Conclusions

The Saber T & T tests are external evaluations, carried out by the Colombian Institute for the Evaluation of Education (Icfes), to students who are about to finish their higher education studies and although many students assume it as a test that responds to the interests of the entities of higher education. Higher education and the state, and not the interests of students. These exams are important because they provide information on the degree of development of the skills of students who are close to completing an academic program, but it should also be quite significant for the same student because it gives him the opportunity to self-evaluate and obtain great benefits such as the condoning of credits with Icfes, among others. It must be important for institutions to review the results of their students because it allows them to make adjustments and make medium-term decisions about their curriculum.

Statistics is a powerful and fundamental tool that allows us to infer the behavior of certain events, it also allows us to obtain accurate information about a group of quantitative data and puts them at the service of decision making.

The Wilcoxon test is non-parametric that contrasts two samples) of the Relationship between a qualitative variable (in the present investigation the Saber Tests) and another quantitative variable

(the Results) and that it makes a comparison contrast of the central tendency, in this case the median. This tool was used to correlate the results of Saber 11 Test and Saber T & T in the Institutions of Higher Education IES. Thus becoming a tool that can help the Faculties evaluate their academic programs in relation to the quality of education that provide and the curriculum they offer and in this way make decisions based on solid foundations generating improvement actions to strengthen the skills that students need to meet the objectives of any institution and achieve educational quality.

References

- [1] Ministry of National Education 2016 *Document of Self-evaluation for the purposes of renewal of the international certification of the national accreditation council* - CNA colombia Bogotá October 2016 pag. 9
- [2] *Dynamics of inclusive education in colombian higher education: process, principles and challenges*
- [3] *Self-evaluation Document for the Purpose of Renewal of the International Certification of the national Accreditation Council* - CNA colombia, p. 13
- [5] <https://www.cna.gov.co/1741/article-187264.html> last access Dec 15 2018
- [6] Minimum Quality Conditions of Academic Programs 2004 Bogotá, colombia, Ministry of National Education Vice Ministry of Higher Education Quality Department
- [7] http://www.tecnar.edu.co/tecnar/historia_ last access Dec 15 2018
- [8] Anderson D R, Dennys J S and Williams 2008 *Statistics for Administration and Economics* 10th Ed Editorial Asv
- [9] Corso S., Jimmy A 2005 *Non-parametric statistics* Editorial UN