Journal of Education and Human Development
December 2016, Vol. 5, No. 4, pp. 157-166
ISSN: 2334-296X (Print), 2334-2978 (Online)

Published by American Research Institute for Policy Development
DOI: 10.15640/jehd.v5n4a15
URL: https://doi.org/10.15640/jehd.v5n4a15

Validation of an Instrument Designed to Measure the Academic Writing and Reading Abilities of Students in the Field of Education

Martha Olivia Ramírez Armenta[1], Ramona Imelda García López[2], Jesús Tánori Quintana[3] & Maricela Urías Murrieta[4]

Abstract

The study had the purpose of determining the metric properties of an instrument that was designed to measure the academic writing and reading abilities of students in the field of education. For this, a quantitative study was carried out, with a correlational reach and a non-probabilistic sample of 222 students. The statistical analysis consisted of: measuring the level of difficulty of each reactive, determining the value expressed in quadratic measures for internal and external adjustment, observed biserial point and discrimination. The obtained results through the aforementioned tests show that the instrument possesses theoretical and empirical subtenant in order to measure academic writing and reading abilities. For the hypothesis tests, descriptive statistic was used to analyze the behavior of said values. The findings show that students lack ability in the field of measured competence.

Key Words: Academic literacy, university students, reliability and validity, academic text, reading, writing

1. Introduction

The formation within academic reading in writing of college students has been proclaimed as one of the highest-interest elements by institutions and scholars of higher education. Said interest is based on how determinant these abilities are in order to reach success in higher education; being that by not dominating them, the individual does not possess the minimum necessary tools required for their formation within this discipline and later, during their duties as a professional (Salazar et al., 2015). Thus, proposals for university formation, such as academic alphabetization; regarded as the ability to use reading and critical and epistemic writing as tools to own and produce knowledge within the specific discipline of formation for students and/or professionals, arise (Guzmán & García, 2015).

The concept of academic alphabetization emerged in the late 90´s in North American Universities; however, it was established with the movement writing across the curriculum, which arose 20 years earlier. Said movement states that in order for students to be better learners and thinkers they must become better readers and writers. Likewise, it established that writing is a tool that allows us to think and transform our own knowledge.

[1] The affiliation of all authors is: Instituto Tecnológico de Sonora, Education Department. Address: 5 de Febrero 818 sur Colonia Centro, Ciudad Obregón, Sonora México.
[2] The affiliation of all authors is: Instituto Tecnológico de Sonora, Education Department. Address: 5 de Febrero 818 sur Colonia Centro, Ciudad Obregón, Sonora México.
[3] The affiliation of all authors is: Instituto Tecnológico de Sonora, Education Department. Address: 5 de Febrero 818 sur Colonia Centro, Ciudad Obregón, Sonora México.
[4] The affiliation of all authors is: Instituto Tecnológico de Sonora, Education Department. Address: 5 de Febrero 818 sur Colonia Centro, Ciudad Obregón, Sonora México.

For this reason, the ability must be developed at the same rate as disciplinary subjects and not during the years previous to a university formation (Sánchez, 2016).

In the field of research there are many disciplines that have been involved in performing tests about the deficiencies presented in the use of the language among university students. The main detected deficiencies are that the students do not read assigned texts, such as the basic bibliography necessary to successfully complete a subject, in a comprehensive manner. In the case of reading practices that the students use orientated to evaluations, the scholar is able to recognize that at multiple times there have existed inadequate or partial interpretations of them (Reyes & Fernández, 2015). Regarding redaction, a significative amount of the papers written by university students lack coherency as well as proper and strong arguments, and the contained information presents plagiarism from the used sources in many cases, whether willing or unwillingly (Booth, Colomb, & Williams, 2008).

In general, the importance of performing tests in the area of university-level language lies on how elemental these abilities are in order to develop critical thinking, autonomy in learning and the ability to solve student´s necessities. In this sense, academic alphabetization emerges as a proposal that seeks answering the current necessities presented by the academic and work environment; just as much in the diversification of access and in the permanent and autonomous formation development (Guzmán & García, 2015).

The objective is to build, validate and obtain the level of reliability of an instrument that measures the academic reading and writing competence of students in the field of education. As well as proving the hypothesis that have to do with the already created and validated dimensions such as the current semester studied by the students, the gender of said students and whether the courses were accredited or not.

2. Theoretical framework

2.1. Evaluation

In the educational environment, the process of evaluation is a complex subject of vital importance that requires a deep interest from every individual involved during the formative process. This is due to the fact that it is during this process that any sort of immediate reality attached an established phenomenon can be observed and analyzed; with the intention of making better decisions to improve the educational process (Consejo Mexicano de Investigación Educativa, 2005). That is to say that the evaluation is not only a mechanism that can be used to give a score, it is neither the simple calculation of the value of any specific learning, nor does it work as a form of control (Ibarra & Rodríguez, 2010); it is something that goes further; since it is also a process that provides useful information about the planning, realization and impact of a determined object, thus allowing to offer a guide in the process of taking decisions (Jiménez, González & Hernández, 2010).

In this sense and following the afore mentioned perspectives, an evaluation must be guided by a model that further solidifies the process; which should provide an answer to multiple key questionings, such as; what is it that will be evaluated, the implicated parts in the evaluation, who is evaluating, how is he or she doing it, what with and with what intentions, among other (Muñiz & Fonseca, 2008). The extent to which the evaluation process tends to the implicated methodological aspects will be the level of reliability that the collected data will have. Because of this, the creation, validation and/or adaptation of said instruments are a fundamental element within the process of evaluation; since analyzing the properties which determined an instrument possess allows us to offer reliable information for the taking of decisions(Ibarra & Rodríguez, 2010).

2.2 Metric properties of the evaluation instruments

As it was mentioned earlier, it is essential that during the process of evaluation the technically appropriate tools are available for its realization. In general, the necessary technical elements are that the instruments possess validity and reliability. These elements consist of performing different analysis of the statistical type to the test with the intention of establishing parameters and determining whether the reactive that make the instrument function properly (Aiken, 2003; González, 2012). In order to obtain the level of validity and reliability of an instrument, the items are analyzed in a quantitative and qualitative manner. In a similar fashion and with the intention of selecting the most adequate reactive from a metric perspective for these tests, the following is taken into account: the index of difficulty, the discrimination index and the differential operation of (Muñiz, Fidalgo, García, Martínez, & Moreno, 2005).

Within the area of reactive analysis there are multiple techniques and mathematical procedures to be able to assess the quality and pertinence of the items. In order to perform an estimate of reliability, strategies or models based in the Classic Test Theory (TCT, for its Spanish acronym) are used. The estimations regarding the validation process are performed by observing the instrument's relation to other instruments and is placed within a scale or parameter (Muñiz, 2010).

This work is based on the theory of the item's answer due to the fact that it is the one with the widest projection within the educational field; it is important to notice that said theory does not contradict the TCT assumptions, but rather proposes additional assumptions that allow us to answer matters that the TCT could simply not (Muñiz & Fonseca, 2008).

2.3 The Theory of an Item's Answer (TRI)

It is a theory that intends to describe, through a mathematical model, the relation given between the subject answering the tests and the question or reactive said individual faces (Instituto Dominicano de Evaluación e Investigación de la Calidad Educativa, 2014). Unlike other theories, TRI allows us to utilize methods that can be judged empirically; which allows the formation of judgement regarding the validity of the obtained estimations (Asún & Zúñiga, 2008). Said estimations do not vary since they are non-dependent of the specific group to which the test is given. Furthermore, the calculation of internal consistency, as a factor of the level of reliability, in this theory allows obtaining measures for errors for each individual or ability level in particular; thus being able to determine how precise the measurement carried out is. On the other hand, the estimations of the items characteristics and individuals within TRI are obtained with the same measurement scale or parameters, thus facilitating the comparison (Asún &Zúñiga, 2008).

Based on this, it can be generalized that the TRI details in an isolated manner the behavior of items and individuals. Furthermore, it assumes that the answer given by the subject depends on the level of ability or knowledge they possesses when evaluated.

2.4 Rasch's Metric Model

This model sustains two assumptions: (a) the range wished to be measured can be represented in a sole dimension in which people and items and the level in which the person with the content are found and the level of difficulty of the item determine the probability that the given answer is correct. In other words, the model, through a mathematical formula, allows us to relate the success probability with differences between the individual's ability and the difficulty of the item (Prieto & Delgado, 2003).

This model is utilized to measure a phenomenon from a series of obtained scores for different items. Namely, it allows us to represent the phenomenon wished to be calculated (Martín, Díaz, Córdoba, & Picquart, 2011). The data with which this model works is dichotomy (wrong or right answers); and allows the description of results with different intensities. The probability that a subject responds correctly to a question does not only depend on the subject's intelligence, but on the difficulty of the item as well (González, 2012).

Within this model, reactive must have certain interpretation criteria: internal adjustment (infit) and external adjustment (outfit). The infit is a value sensible to irregular answer patterns affected by unexpected answers and its value oscillates around 0.7 and 1.3 in order to be considered in a situation of productive measurement (it is considered a perfect situation when the value comes close to 1.0. Quantified values under 1.0 are considered plentiful with determinism; whilst values over 1.30 are considered random; namely, that the subject has responded randomly to the test. The outfit value determines atypical cases in non-expected response patterns. It is determined by averaging the standardized quadratic deviations between the observed and the expected performance. It detects disadjustments in deviations far from the item's measure zone. Thus, it must be kept within a range of 0.7 to 1.3 (Bazán, 2000).

The biserial value point is an index of internal validity that allows to determine the grade of validity and the item's pertinence; this to say, it establishes whether the question measures what the test proposes to measure (Chávez & Saade, 2010). The measure *Measure* indicates the level of difficulty of each question. If the question has a difficulty index close to 0 or 1 it should be modified or completely discarded since it is not giving information about the difference of ability among subjects; namely, not discriminating.

What is pretended to be measured with the level of discrimination of a reactive is the grade in which the item is allowed to make a difference between the subjects that obtained a total high score on the test and those that obtained a low score. Said discrimination measure is associated to the subjects that obtained a high score; whilst the negative value reveals that the question discriminates by favoring the group of subjects hat obtained a low score (Chávez & Saade, 2010).

On the other hand, and with the intention of sustaining the results with the validation of the instrument; different hypothesis tests are performed. Specifically, the intention is to observe whether the gender of the subjects, having unaccredited courses and the year in which they took the course had a relation with any of the dimensions conforming said instrument.

3. Method

A quantitative study was performed due to the fact that it was determined under a dominance model, the interaction of the item with the student in variables such as: cohesion and coherence, academic vocabulary, academic genres, graphic reading and text comprehension. In the same manner, it is descriptive since it described the behavior of the mentioned values in an exploratory manner. Finally, it has a transectional reach since values were measured in an independent manner and at one sole moment (Bisquerra, 2009).

3.1 Participants

The study was performed with a non-probabilistic sample of 222 students of the Education Science Bachelor Degree course of a University in the south of Sonora. Said sample consisted of 64 men and 158 women, it is important to note that the apparent disproportion between the samples is due to the nature of the disciplinary field; since the population belongs, in its majority, to the female gender. The average age of the participants was 21-years-old.

3.2 Instrument

The used instrument for the collection of data was a test; this type of instrument was resorted to due to the fact that its application was appropriate since it allows the evaluation of the actual level of dominance that the students possesses. Said test was confirmed by six areas: age, gender, occupation (whether they worked or not and in case they did, the schedule devoted to their occupation), semester they attended, amount of credited and unaccredited courses up to the point of this examination and previous high-school studies (public or private institution). The five remaining areas evaluated the variables shown in charts 1 and 2.

Chart 1: Instrument's specifics chart

Variables	Test Procedures	Items	Level
Cohesion and coherence	Disordered paragraph	7	Application
Academic vocabulary	Construct Dominance	8	Comprehension
Academic genres	Type of text identification	8	Comprehension
Graphic reading	Bar graph interpretation	7	Comprehension
Text comprehension	Topic Identification	7	Comprehension

Chart 2: Variable specifics chart

Variable	Definition
Cohesion and coherence	Fundamental properties regarding text production and comprehension. Coherence allows for an adequate organization and distribution of the given information. Cohesion establishes the proper lexical and grammatical conexions that will give way to comprehension.
Academic vocabulary	It pertains to the concepts and phrases used within a determined discipline, which are fundamental for students and scholars to comprehend and explain the constructs taught at university.
Academic genres	These are the texts elaborated within the academy with the intention of broadcasting scientific knowledge
Graphic reading	Correct representation and interpretation of graphics instruments.
Text comprehension	Ability to understand what is read. It has two related aspects: 1. Structuring a semantics representation and 2. Accomplishing coherence relations between phrases

3.3 Procedure

For the application of the instrument, the approval of the necessary authorities was sought after. When said contact was established it was explained to them the intention of the instrument, the characteristics of it and the utility that the collected data would have for the program and the institution The authorities provided the data that allowed the facilitation of the application, such as: the amount of students enrolled on each course (with the intention that the application was performed during hours where the highest percentage of students possible were present); the names of the scholars (to be able to obtain permission to enter the classrooms) and finally; they had the disposition to send notice to the scholars intending for them to allow the application of said instrument, which further facilitated the entrance to the classrooms.

The used space for the application was the researched university classrooms. The visit would be initiated with the presentation of the scholar; afterwards, the objectives and importance of the study would be explained to the students; the reliability and privacy of their results, which would be reported in a generalized manner, was spoken about; as well as thanking them for their participation. The duration of the application was estimated to an hour and the majority of the students agreed to answer the instrument.

3.4 Statistical Analysis Procedure

The instrument was subjected to the necessary analysis to determine psychometric properties of each reactive, the unidimensionality of the scale and the reliability. The aforementioned was performed using the Item Response Theory (TRI), specifically with the Rasch model of measurement using the Winsteps software (González, 2008).

For the result analysis, the statistical analysis software Statistical Package for the Social Sciences (SPSS) was used. Specifically, descriptive analyses were performed to observe the dependent variable distribution. For this, Pearson, Student T de student, One-Way Anova and the non-parametric Mann-Whitney tests were used. The reliability estimation of the instrument was performed through the KR- 20 coefficient.

4. Results and discussion

37 reactives were subjected to the necessary statistical analysis in order to determine the metric properties of the dominance instrument. These consisted of: measuring the difficulty of the reactive, determining the adjustment values expressed in quadratic measures for internal and external adjustment, observed biserial point and discrimination.

In order for the reactive to be considered as a part of the instrument, its discrimination index should lie within a scale of -2.5 a 2.5. The accepted infit and outfit values for each reactive were from 0.7 to 1.3. The difficulty index should approach 50% (percentage considered to be the middle level of difficulty) and a biserial point correlation with the higher than .20 scales. Nevertheless, the most important criteria for an item to be kept were the in fit and outfit values due to the fact they demonstrate the item-subject relation (Chávez & Saade, 2010).

Under these criteria, 31 reactive that were within the acceptable levels were kept. The obtained results throughout the different metric analysis performed are shown in chart 3. In these, the properties of items through the Rasch model were analyzed. The reliability test was performed with the KR – 20 coefficients. As it is shown, the reactive´ s infit and outfit values that conform the final test lied within the acceptable limits (0.7 to 1.3). In the same manner, these reactives maintained a discrimination index approaching one. The difficulty was kept close to the required 50%and the correlation with the scale was kept above .20, which allows to determine that the items correspond to the construct‘s measurement. Furthermore, the initial model gave a reliability result of -95. The items conforming the test work towards a unidimentional variable, but conglomerate in groups with the intention of finding about which dimensions are sensible to certain contextual variables. The created dimensions were: cohesion and coherence, academic vocabulary, academic genres, graphic reading and text comprehension.

Chart 3: Metric analysis set results

Reactive Total: 31		**KR- 20 Coefficient:** .95				
# REACTIVE	MEASURE	ERROR	INFIT MSQ	OUTFIT MSQ	PBSE	DISCRIM
Item 21	62.02	1.92	1.01	1	0.11	0.99
Item 20	60.64	1.84	0.96	0.88	0.21	1.06
Item 29	59.36	1.78	0.98	0.95	0.16	1.03
Item 31	59.36	1.78	1.03	1.13	0.05	0.94
Item 4	59.05	1.77	1.01	1.06	0.09	0.97
Item 6	58.75	1.75	1	1	0.13	1
Item 28	58.46	1.74	1.01	0.99	0.11	0.99
Item 27	56.76	1.67	1.02	1.01	0.1	0.97
Item 37	55.7	1.63	0.96	1.01	0.19	1.05
Item 24	53.47	1.56	0.96	0.93	0.22	1.11
Item 5	52.53	1.54	1.02	1.03	0.11	0.95
Item 30	52.08	1.53	0.93	0.93	0.26	1.18
Item 2	51.4	1.51	1	1	0.14	0.99
Item 22	51.18	1.51	1.03	1.04	0.09	0.89
Item 3	50.96	1.51	1.05	1.05	0.07	0.84
Item 34	50.52	1.5	1.04	1.11	0.07	0.82
Item 38	50.52	1.5	1	1	0.14	0.98
Item 17	50.09	1.49	1.01	1.14	0.13	0.93
Item 33	49.87	1.49	1.01	1.01	0.13	0.95
Item 18	49.02	1.47	1.05	1.05	0.07	0.79
Item 23	47.77	1.46	0.99	0.99	0.16	1.04
Item 15	47.36	1.45	1	0.98	0.16	1.04
Item 32	44.72	1.44	0.98	0.97	0.18	1.13
Item 19	44.32	1.44	0.95	0.92	0.24	1.39
Item 10	42.51	1.44	0.97	0.95	0.21	1.22
Item 25	41.5	1.45	0.92	0.92	0.29	1.48
Item 13	41.29	1.45	1.01	1.01	0.12	0.93
Item 11	39.44	1.47	1	1.04	0.13	0.94
Item 12	38.17	1.49	1.04	1.06	0.07	0.82
Item 26	37.3	1.5	1.01	0.99	0.13	0.97

4.1 Attributive variable relation with the dimensions of academic reading and writing

The intention was to show the effects of the application of said instrument; namely, the obtained results with this test are shown. The first step after the validation of the instrument was searching for the correlation between socio demographic data and the dimensions contained in the test (see chart 4). In this procedure it was found that there is a correlation between the semester the students attended at the time of the application and the dimension of academic vocabulary. This is to say, it was found that a set of grouped students per semester presented a difference in regard to the others. With this tendency more specific tests were performed with the intention of finding which was the group making the difference?

Chart 4: Correlation between socio demographic data and the different variables

Variables	1	2	3	4	5	6	7	8
1. Current semester	-							
2. High school studies (public or private)	.078	.085	-					
3. Cohesion and coherence	.121	.039	-.065	-				
4. Academic vocabulary	.229**	.033	-.028	.122	-			
5. Academic genres	.071	-.013	-.044	.031	.160*	-		
6. Graphics reading	.079	.059	.033	.114	.093	.127	-	
7. Comprehension	.076	-.001	-.060	.155*	.145*	.101	.031	-

The correlation is significant at the 0, 01** level. The correlation is significant at the 0, 05* level.

Afterwards, the T test was utilized for independent samples with the intention of identifying significative differences between the gender of the subjects and the dimensions of the instrument. As shown in chart 5, differences in the dimension of graphic reading were found.

Chart 5: Differences between the gender of the subjects and the academic alphabetization variables

Variable	Male		Female		t(220)	*p*	Cohen's d
	M	SE	M	SE			
Cohesion and coherence	28.75	24.656	27.97	21.048	0.23	.813	0.03
Academic vocabulary	55.72	23.98	56.75	23.19	-0.29	.769	-0.03
Academic genres	29.49	18.68	34.49	21.32	-1.63	.103	0.01
Graphics reading	40.17	23.31	31.28	20.78	2.78	.006	0.37
Comprehension	38.75	23.60	37.72	39.42	0.19	.846	0.11
General score	38.35	37.72	12.48	13.33	0.32	.746	0.09

In this dimension, male students obtained a higher score than female students; it can be inferred that said inequality is due to the different reasoning abilities possessed by men and women. The results found through this process agree with those obtained during the investigation about the difference between the cognitive abilities and academic performance during university studies between men and women. In said study, men obtained better results in calculus and abstract reasoning tests, which are necessary abilities for graphic reading; since they are constituted by logical and mechanic thought that allows making inferences to understand the result or value of something specific through a numerical expression (Echavarri, Godoy, & Olaz, 2007).

As it is shown in the chart 5, the potency of difference through the Cohen's test (Cohen, 1992) tends to be fairly potent. The aforementioned, due to the fact that it complies with the criteria to state that the difference between mediums in men and women in the dimension of graphic reading is not false (Cárdenas & Arancibia, 2014).

Subsequently, the ANOVA test was used to identify whether any difference between the year the students course and any dimension contained within the instrument existed. With this test it was proven that there are differences regarding the general score and the dimension of academic vocabulary, with one of the semester groups (see chart 6).

Chart 6: ANOVA variance analysis for the different dimensions of linguistic abilities and the subject's current semester variable

Variable	1st. year		2nd. year		3rd. year		4th. year		ANOVA		n^2
	M	SE	M	SE	M	SE	M	SE	F (3,222)	*p*	
1	24.79	20.96	28.78	24.51	29.17	23.68	30.97	20.58	.928	.428	0.10
2	49.06	22.16	54.47	24.72	62.84	22.86	61.29	22.13	4.781	.003	0.21
3	29.92	20.64	35.36	19.54	33.85	18.58	34.47	22.93	.831	.478	0.08
4	30.78	18.64	35.54	22.83	32.14	21.14	37.55	24.84	1.243	.295	0.12
5	33.80	26.53	47.31	63.83	35.00	25.92	39.03	22.15	1.409	.241	0.20
6	33.62	12.30	39.96	15.23	38.50	12.13	40.99	12.04	4.256	.006	0.13

1= Cohesion and coherence; **2**= Academic vocabulary; **3**= Academic genres; **4**= Graphics reading; **5**= Comprehension; **6**= General score
1st. year = first and second semester; **2nd year** = third and fourth semester; **3rd year**=fifth and sixth semester; **4th year**=seventh and eighth semester.
Post hoc: 1= 1st. year, 2= 2nd year, 3= 3rd. year, 4= 4th. year.

With the intention of determining groups where differences are presented the Post Hoc test was utilized, specifically the Bonferroni test (González, 2008). With this test it was found that there are differences between the first year groups with those in the fourth year. As it is shown in Chart 6, the differences lie in the fact that fourth-year students obtained better results than students that attended the first year; both in the dimension of the general test score and the academic vocabulary dimension. These results show that the one making the difference is the first-year group, since they obtained lower scores than the students about to graduate.

Finally, a test was performed to determine the size of the effect; this is to say that the n^2 of the four-group comparison with different variables was calculated. With this test it was stated that the two dimensions that were significant turned out to be true, thus rejecting a false positive and discarding a Type 1 error (Cárdenas & Arancibia, 2014).

In general, the reached average between groups is not broadly different; and although it seems to be that the approach that the university provides with literate practices causes a small improvement, this is not enough. The aforementioned, due to the fact that none of the groups obtained passing indexes; but the means among groups showed that the students attending their fourth year raised the obtained average score in these two dimensions. This is due to the fact that the approach of the youth towards scientific texts improves their notions in the knowledge of language; however, due to the fact that there is not an established commitment as such to form them as critical and autonomous writers and readers the progress is not as substantial. What is worrying is that said improvement is not necessary for the required level of a professional (who should be able to dominate these abilities entirely). Specifically, these results make up important data, since this degree has the education of other people as a field of work. As far as the academic vocabulary dimension, the progress set a greater difference; what allows us to reckon that throughout the course students acquire the basic concepts of the science they are being formed in and are capable of defining them.

For to end, with the Mann-Whitney test, differences in academic writing and reading were identified among students that had non-accredited subjects throughout the coursed semesters against the students that had successfully completed every course. Regarding this, differences in vocabulary dimension were found. It was agreed to utilize this non-parametric test due to the fact that data did not maintain the same normality parameters as in the other variables; however, since there is an interest in studying its behavior, this type of test was resorted to (Chávez & Saade, 2010). As it can be seen in chart 7 the obtained results were that the subjects with non-accredited subjects had higher scores than the subjects who accredited all subjects. Namely, the average high range is found among students who have not accredited a course, being this significance 0.40.

Chart 7: Descriptive statistic in academic vocabulary dimension and accredited and non-accredited subjects variable

Ranges				Test Statistics		
Total score in academic vocabulary dimension	Non-accredited subjects	n	Averange range	Mann-Whitney U	Z	*p*
	No	80	69.28	2302	-2.05	.040
	Yes	71	83.58			
	Total	151				

According to these results the students that repeat classes have a higher index of appropriation of concepts and its definitions. This result can be associated to the studies performed on the acquisition of vocabulary in a second language. In said study it was found that repeating and relating the concepts with different contexts (regardless of it being for exemplification or explanation) the student relates them by creating memory nets, which allows him or her to reproduce said concepts voluntarily in a written as well as an oral manner; thus achieving to hold on to them for a longer period (Moreno, 2013).

5. Conclusions

The previously obtained results suggest that the performed test contains the metric properties that allow using it as an instrument to measure the academic writing and reading abilities of students in the field of education. The dimensions of the instrument are empirically sustainable to be used as tools to evaluate this construct; since it has validity evidence and reliability.

The results obtained by the students provide concrete evidence of the discernments that result from interest in the study of the language or scholar labor; since it blatantly proves the subpar linguistic knowledge of the students and shows how the mastery of it is far from being what is expected from a professional These results manifest the inconsistencies of the educational system; since, on one hand, the intention is that undergraduate students, graduate students and even professionals comprehend and produce complex texts in the field of science without having the necessary competences to manage it. Furthermore, it is required that they build arguments, ponder critically and communicate through discursive structures that are appropriate to each discipline without giving them the required knowledge in order for it to happen.

The process of improving the professionals and student's communicative competence represents a challenge that higher education institutions must arise to, simply because of the transversality this represents.

References

Aiken, L. (2003). *Test Psicológicos y Evaluación (11ed).* México: PEARSON.

Asún, R., & Zúñiga, C. (2008). Ventajas de los Modelos Politómicos de Teoría de Respuesta al ítem en la Medición de Actitudes Sociales: El Análisis de un Caso. *Psykhe*, *17*(2), 103-115. Available http://www.scielo.cl/scielo.php?pid=S0718-22282008000200009&script=sci_arttext

Bazán, J. (2000). *Evaluación psicométrica de las preguntas y pruebas crecer 96.* Ministerio de Educación. Available http://www2.minedu.gob.pe/umc/admin/images/menanexos/menanexos_206.pdf

Bisquerra, R. (2009). *Metodología de la investigación educativa*. Madrid: La Muralla.

Booth, W., Colomb, G., & Williams, J. (2008). *Cómo convertirse en un hábil investigador.* Barcelona: Gedisa.

Cárdenas, M., & Arancibia, M. (2014). Potencia estadística y cálculo del tamaño del efecto en G*Power: complementos a las pruebas de significación estadística y su aplicación en psicología. *Salud y Sociedad*, *5*(2), 210-224. Available http://www.redalyc.org/articulo.oa?id=439742475006

Chávez, C., & Saade, A. (2010). *Procedimientos básicos para el análisis de reactivos.* México: Centro Nacional de Evaluación para la Educación Superior.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 12* (5), 155-159.

Consejo Mexicano de Investigación Educativa. (2005). La evaluación educativa. *Revista Mexicana de Investigación Educativa, 10*(27), 1257-1258. Available http://www.redalyc.org/articulo.oa?id=14002717

Echavarri, M., Godoy, J., & Olaz, F. (2007). Diferencias de género en habilidades cognitivas y rendimiento académico en estudiantes universitarios. *Universitas Psychologica*, *6* (2), 319-329. Available http://www.redalyc.org/articulo.oa?id=64760211

González-Montesinos, M. (2008). *El Análisis de reactivos con el Modelo Rasch: Manual Técnico*. México: Universidad de Sonora: Instituto Nacional para la Evaluación de la Educación.

González-Montesinos, M. (2012). *Modelos de medición. Modelación de Ecuaciones Estructurales*. México: Centro de Investigación en Alimentación y Desarrollo, A.C

Guzmán, F., & García, E. (2015). La alfabetización académica en la Universidad. Un estudio predictivo. *RELIEVE, 21* (1),1-16. Available http://www.redalyc.org/html/916/91641631006/

Guzmán, F., & García, E. (2015). La evaluación de la alfabetización académica. *RELIEVE*, *21*, 1-16. Available http://www.redalyc.org/html/916/91641631006/

Ibarra, M., & Rodríguez, G. (2010). Los procedimientos de evaluación como elementos de desarrollo de la función orientadora en la universidad, *REOP*, *21(*2), 443-461. Available http://www2.uned.es/reop/pdfs/2010/21-2%20-%20Soledad%20Ibarra.pdf

Instituto Dominicano de Evaluación e Investigación de la Calidad Educativa. (2014). *Introducción a la Teoría de Respuestas al Ítem*. República dominicana: Ministerio de Educación.

Jiménez, Y., González, M., & Hernández, J. (2010). Modelo 360° para la evaluación por competencias (enseñanza-aprendizaje). *Innovación Educativa, 10*(53), 43-53. Available http://www.redalyc.org/articulo.oa?id=179420770003

Martín, N., Díaz, C. Córdoba, G. & Picquart, M. (2011). Calibración de una prueba química por el modelo de Rasch. *Revista Electrónica de Investigación Educativa, 13*(2), 132-148. Available http://redie.uabc.mx/index.php/redie/article/view/288

Moreno, J. (2013). Recursos multimedia y mnemotecnia para la adquisición de vocabulario en inglés, *Revista virtual Universidad Católica del Norte,* (38), 112-126. Available http://www.redalyc.org/pdf/1942/194225730009.pdf

Muñiz, J. (2010). Las teorías de los test: Teoría clásica y teoría de respuesta a los ítems. *Papeles del psicólogo 31*(1), 57-66. Available http://www.papelesdelpsicologo.es/pdf/1796.pdf

Muñiz, J., & Fonseca, E. (2008). Construcción de instrumentos de medida para la evaluación universitaria. *Revista de Investigación en Educación,* (5), 13-25. Available http://reined.webs.uvigo.es/ojs/index.php/reined/article/viewFile/40/35

Muñiz, J., Fidalgo, A., García, E., Martínez, R., & Moreno, R. (2005). *Análisis de los ítems.* Madrid: La muralla.

Prieto, G., & Delgado, A. (2003). Análisis de un test mediante el modelo de Rasch. *Psichothema 15* (1), 94-100.

Reyes, S., & Fernández, J. (2015). La escritura académica en la formación en línea de docentes de escuelas públicas mexicanas. *Revista Electrónica de Investigación Educativa*, *17*(2), 1-15. Available http://redie.uabc.mx/redie/article/view/412

Salazar, A., Sevilla, O., González, B., Mendoza, C., Echeverri, A., Quecán, D., Pardo, L., Angulo M., Silva, J., & Lozano, M. (2015). Lectura y escritura en la universidad: contribución para reconstruir una historia. *Magis Revista Internacional de Investigación en Educación, 8* (16), 51-70. Available http://www.redalyc.org/pdf/2810/281042327004.pdf

Sánchez, S. (2016). El letrismo académico a través de las asignaturas en las universidades públicas estatales de México. *Revista Iberoamericana de Educación Superior, VII* (19), 131-151. Available https://ries.universia.net/article/viewFile/1137/1869.