

Walter Araya
walter.arayagarita@ucr.ac.cr
José Alejandro Fallas
jose.fallasgordinez@ucr.ac.cr

Original scientific paper
UDC: 371.279.7
DOI: 10.5937/IstrPed2401001A

=====

UNIVERSITY OF COSTA RICA'S ENGLISH DIAGNOSTIC TEST 2022: EVIDENCE OF VALIDITY AND RELIABILITY

Abstract: The English Diagnostic Test aims to assess reading comprehension skills for first-year students at the University of Costa Rica. In 2022, this test consisted of four instruments with 55 items. Instruments were based on an academic reading, following the criteria B2+ or C1 level according to Common European Framework of Reference for Languages. This paper explores the evidence of validity and reliability from 1732 first-year students virtually assessed in March of 2022, using the platform of the Programa de Evaluación en Lenguas Extranjeras (PELEX). Results from this study demonstrated the unidimensionality of the four instruments. Cronbach's alphas for the four instruments ranged from 0.79 to 0.84. Evidence collected in this study supports the work done by the School on Modern Languages in the development of valid and reliable instruments for the English Diagnostic Test, as part of the current needs of new students for undergraduate programs at the University of Costa Rica.

Keywords: reading comprehension, assessment, validity, reliability

PRUEBA DE DIAGNÓSTICO DE INGLÉS DE LA UNIVERSIDAD DE COSTA RICA 2022: EVIDENCIAS DE VALIDEZ Y CONFIABILIDAD

Resumen: El examen de diagnóstico de inglés tiene como objetivo evaluar las habilidades de comprensión lectora, para la población de primer ingreso de la Universidad de Costa Rica. En el año 2022, la prueba consistió en cuatro instrumentos conformados por 55 ítems cada uno. Cada instrumento fue elaborado con base en una lectura académica, las cuales debían responder al criterio de B2+ o C1 según el Marco Común Europeo de Referencia para las lenguas. Este artículo explora evidencias de validez y fiabilidad correspondiente a la evaluación virtual de 1732 estudiantes de nuevo ingreso en marzo de 2022, mediante la plataforma del Programa de Evaluación en Lenguas Extranjeras (PELEX). Los resultados de este estudio demostraron la unidimensionalidad de los cuatro instrumentos. Los coeficientes de Cronbach de estos cuatro instrumentos se encontraron en el intervalo de 0,79 y 0,84. Las evidencias de recolectadas en este estudio apoyan el trabajo de la Escuela de Lenguas Modernas, en el desarrollo de instrumentos válidos y fiables para el examen de diagnóstico de inglés, como parte de las necesidades actuales de los estudiantes de nuevo ingreso, en los programas de grado de la Universidad de Costa Rica.

Palabras Claves: comprensión lectora, evaluación, validación, confiabilidad

Introduction

Language testing should rely on the principles of equality and social justice, mainly for those tests that possess direct high consequences for examinees. High-stake language tests require the use of statistical techniques for collecting evidence of internal validity and reliability. This evidence can assure examinees, researchers, test developers and stakeholders that instruments are valid and reliable for measuring the language ability or abilities. Test analyses for validation and reliability guide decision-making of tests developers, but they also provide trustworthiness of the instruments.

Construct is the term used for a general concept, ability, characteristic or latent trait, which works as the base and reason of measurement. In this case, the English Diagnostic Test aims to measure the construct of reading comprehension skill in English as a second language. Programa de Evaluación en Lenguas Extranjeras (2022) defines reading comprehension as:

... the use of a variety of reading strategies that would allow users to participate actively and logically in the construction of reading material, at the same time, they can understand and analyze such texts by taking a subjective position considering the author's perspective. (p. 1)

Validity is the process of assuring that an instrument genuinely measures the latent trait, defined as part of the construct. Martinez, Hernandez, & Hernandez (2014) mention different divisions for validation though they can be summarized into validation of the construct. This paper focuses on showing evidence of internal validity of the construct, throughout the application of *Factor Analyses* (FA). These techniques provide evidence of the internal structure of a test.

Martinez, et al. (2014) explain the main difference between both methods. The *Exploratory Factor Analysis* (EFA) does not assume a priori number or relationships among factors. Factor errors are independent, and the main purpose is to find a structure of dimensions or latent constructs, within the items. Contrarily, the *Confirmatory Factor Analysis* (CFA) tests the hypothesis of a priori structure in the form of a model. It allows to establish correlation between factor errors. Factor Analyses generates information of the dimensionality of a test. In this case, each instrument bases on the premise of one dimension of measurement, reading comprehension skills.

Another key aspect for an instrument is the consistency of its measurement. Reliability implies that the instruments used are consistent if reproducing its attempts to measure the latent trait (ALTE, 2011; Mair, 2018). Strong evidence of internal consistency supports higher levels of reliability. Some measurements of internal consistency are Cronbach's alpha (1951) and Spearman-Brown's (1910) split-half method. Both methods estimate the internal consistency of any instrument. Desirably, high impact tests should demonstrate high standards of validity and reliability.

High-stake tests require a close examination of evidence for validation and reliability. Some scholars have used the *Classical Tests Theory* (Araya, 2021; Green, 2019) and *Item Response Theory* approaches (Cerdas-Nunez & Montero-Rojas, 2017; Solorzano-Salas & Montero-Rojas, 2011; Zamora-Araya, 2015) to analyze psychometric properties of high-stake instruments in Costa Rica. This paper addresses only the Classical Tests Theory to present evidence of validity and reliability in four instruments for measuring reading comprehension skills. This study explores the validity and reliability process of the diagnostic test (EDI for its acronym in Spanish), applied in 2022 to first-year students at the University of Costa Rica.

Theoretical Foundations

Reading comprehension in different texts is one of the macro skills included in the teaching and learning processes of a foreign language. This skill is one of the most important to be used at the university level because it facilitates the understanding of the most updated bibliography to which

students have access in the higher education system. They should not only read general texts but also texts from different areas such as the arts, social, economic, medical, among others.

Given this scenario, it could be inferred that if students have difficulty understanding a text in a second language, their academic enrichment would be affected, and this situation is detrimental to their academic performance. According to Brizuela, Perez and Rojas (2020), pre-university texts have different discursive characteristics from university texts: in the latter, the use of technical lexis, complex grammar, the variety of superstructures and the demands in the interaction of the contents require a greater development of reading comprehension skills and, therefore, the strategies to better understand the texts should be taught in English language courses, particularly when developing reading comprehension skills. The aforementioned authors also state that the purpose of university education is the individual, social, and professional development of students; consequently, the importance of reading is not limited to the academic environment.

In the current context in which we are living, where information is abundant and requires a greater capacity of analysis of the reader to identify its relevance, veracity, and other aspects. The ability of reading comprehension acquires greater relevance to ensure the quality of students' work at college and therefore when they enter the labor market.

At the level of the University of Costa Rica, the School of Modern Languages offers the LM-1030 Reading Strategies in English I course. This course is designed within a schematic and transactional framework, aimed at helping students use a variety of reading strategies that allow them to participate in the active and logical construction of written text in English. Upon completion of the course, students will be able to actively interact with texts of diverse content and rhetorical patterns by applying appropriate reading skills to understand, analyze the content of English texts, and take a position on the author's perspective.

Students who enroll and pass LM-1030, according to course objectives, are able to apply cultural and formal schemas to identify their genre; make predictions about their content and check them; guess unknown words from context; recognize when and how to use the dictionary; identify rhetorical patterns; punctuate main ideas; recognize details or secondary ideas; and respond critically to a text by evaluating the author's perspective, making inferences from explicit or implicit information, identifying facts and opinions, discriminating ideas (main and secondary), and making judgments about what is read.

It is important to mention that there is not as much research on the use of reading strategies and reading comprehension at the higher education level as there is for another population such as children (Landi, 2010). This situation has contributed to the fact that there are currently more tests to measure low-level skills (word recognition, reading speed, phonological awareness, etc.) than those available to measure high-level cognitive processes associated with the comprehension of complex academic texts (Brizuela, Perez, & Rojas, 2020).

These authors also conducted a study in which they mention that in the Spanish-speaking environment, there are approximately forty instruments oriented to the measurement of reading comprehension for different ages, with response formats of all kinds and designed for educational, clinical, selection purposes, etc. Likewise, there are various techniques that have been used to measure reading comprehension, such as evocation or free recall, probing questions, open response questions and selected response questionnaires (Perez, 2005). The reading comprehension test developed by Dr. Violeta Tapia in 1977 deserves special mention (Lopez, 2010), as it is one of the few that has been designed for the university population. This test measures the following skills: identification of information about specific facts, definition of the meaning of words, identification of the central idea of the text, interpretation of facts, inference about the author, inference about the content of the fragment and labeling (Lopez, 2010). Another test that

deserves a separate mention is CompLec, which was developed by Llorens et al. (2011) and has adolescents as its target population.

In the non-Spanish-speaking world, there is a wide variety of standardized reading comprehension tests, such as the TOEFL or one of the subtests of the SAT. There are other tests of more restricted application, such as the Nelson Denny Reading Test or the Woodcock-Johnson III Diagnostic Reading Battery (there is also a Spanish version, the Woodcock-Munoz). There are also educational tests that include a reading comprehension component and are designed to evaluate this skill in relation to the skills acquired through formal education, such as PIRLS (Progress in International Reading Literacy Study), PISA (Programme for International Student Assessment), PIAAC (Programme for the International Assessment of Adult Competencies), EECL (European Language Proficiency Study), NEPS (National Educational Panel Study), among others. For a detailed description of these types of tests, see the work of Reyes, Castillo, Zuniga and Llarena (2009). Given this lack of reading comprehension tests in Spanish specifically oriented to university students, it is necessary to develop an instrument to identify the possible weaknesses of first-time students entering higher education. In addition to the lack of tests oriented to university students, it is important to add that most of them have been designed without an underlying measurement model, so the following study presents some initial results of the validation of a standardized test for university students entering higher education.

The complexity of reading comprehension is evident from the various theoretical models that describe the cognitive and linguistic processes that underlie it. Some models emphasize the mental representations that readers construct as a result of the comprehension of words, sentences and their respective relationships within a text. Other models focus on the acquisition of the skills required to understand a text. Despite the fact that the models specialize in different aspects of reading comprehension, they have in common the notion that reading comprehension involves the construction of a coherent mental representation in the reader's memory (Kintsch, 2013; Kendeou, van den Broek, Helder, & Karlsson, 2014). In other words, in reading comprehension, a coherent representation must be constructed in memory, based on inferences about the relationships between the internal elements of the text, as well as between these and the reader's knowledge of the topic; all this must be executed under the constraints imposed by the reader's working memory capacity and attentional resources in accordance with the reader's coherence standards (Van den Broek, 2012).

For a diachronic perspective on existing models of reading comprehension, the work of Alexander and Fox (2013) can be consulted. From a synchronic approach, the text by McNamara and Magliano (2009) gives an overview of the most important models regarding the study of this construct. Finally, of obligatory consultation is the text edited by Alvermann, Unrau and Ruddell (2013), in which the psychological, educational and social aspects that make up the research and teaching of reading are extensively exposed. Currently, interactive models, in which comprehension is the product of bottom-up and top-down processes, are considered the best for understanding individual differences in reading comprehension (Leighton & Gierl, 2011).

The model followed in the type of reading comprehension used for the application of the test subject to study in this project is based on the contents of the LM-1030 Reading Strategies in English I course. This model is more interactive than traditional and gives great importance to the role played by previous knowledge and experiences in text comprehension. This concept was introduced by the French psychologist Barlett in 1932 and is known as schema (Murillo, 2008). In other words, readers use their experiences and knowledge to have a clearer concept and ideas about what is being analyzed.

Barlett, cited in Sanchez (2002), states that people could read or hear the same story in a different way. He considers that these differences are affected by time, i.e., the longer some incident had occurred (the greater the experience of the readers), the more elaborations are given to a reading.

From this assertion, one could conclude that second language readers will understand a section if they possess the appropriate formal and cultural schemas that are reflected in a text or if these readers have the ability to adapt new information to the schemas they already possess.

However, Anderson (1999) argues that reading in a second language is also a language problem. This situation is accentuated in readers with a poor level in the first language. For Rosenblatt (1996), our brain is a natural processor of text, which goes beyond printed material: it understands the world around us. From this process, reading becomes a natural act.

The reader must have a construct to understand from word recognition to the construction of vocabulary and contextual cues to the construction of meaning. The construction of meaning should be given from what Makuc (2008) proposes based on Krashen (1989); a text that demands an L + 1, which leads to a greater cognitive effort on the part of the reader.

Readers can use one of the many strategies developed for reading comprehension, which aims to identify rhetorical patterns within the text. For Anderson (1999), rhetorical patterns in written discourse represent the surface reflection of the cognitive operations that develop in an individual's mind as part of his or her theory of the world. Readers can use one of the many strategies developed for reading comprehension, which aims to identify rhetorical patterns within the text. For Anderson (1999), rhetorical patterns in written discourse represent the surface reflection of the cognitive operations that develop in an individual's mind as part of his or her theory of the world.

L. M. Rosenblatt's (1997) aesthetic transaction theory is concerned with the reader's response rather than with what happens at the cognitive level. She sees response as an event, the product of an encounter between two parties. The reader is the most important; he or she is the essential component in this event, for his or her experience with the text becomes a unique act in time and place.

Sanchez Salazar (2002) states that the benefits of reader response as an approach to teaching reading have been demonstrated time and again through empirical research. She adds that a novice reader who is subjected to vital interaction with the text will develop analytical skills that will result in more critical and mature thinking. She concludes that reading focused within a transactional framework has a direct effect on the written compositions of school children from early stages. The aesthetic response is reflected not only in the maturity of ideas but also in their structuring.

Regarding reading comprehension, it can be mentioned that in the field of assessment it is a multifaceted construct, and tests always constitute a finite sample of all possible items that could measure this construct, no reading comprehension test by itself will allow measuring reading comprehension in its entirety, but only partial aspects of it (Van den Broek, 2012).

Changes in the techniques used for the measurement of language-related constructs have been shaped by the contributions of psycholinguistics, sociolinguistics, discourse analysis, and research on second language teaching and acquisition, among others. Over the years, the measurement of skills related to language comprehension and production has included from the more "indirect" strategies (i.e., selection items or incomplete text completion) to the more "direct" ones (i.e., realistic situations of language use and comprehension), realistic situations of language use and comprehension), which has enriched the field of language measurement with a wide variety of types of indicators designed based on objectives and theoretical approaches that are sometimes incompatible (Bachman, 2007) and often not explicit (Leslie & Caldwell, 2009; Perfetti & Adlof,

2012). Likewise, in the development of these instruments, reading comprehension has been conceptualized as information processing, as activation or as a skill (Svetina, Gorin, & Tatsuoka, 2011).

Tests have been used in several countries to measure what, from various theoretical approaches, is considered important for a reader to develop when facing a text (Reyes et al., 2009). Similarly, as evidenced in one of the first studies on reading comprehension tests (Davis, 1944), since that time there was already interest in determining what skills were required to answer the typical questions of this type of measurement instruments.

It is important to note that not all so-called "reading comprehension" tests measure this construct in the same way (Svetina et al., 2011), as each of them emphasizes different facets of it. Because reading comprehension is a complex construct that involves the coordination of several cognitive processes (O'Reilly, Weeks, Sabatini, Halderman, & Steinberg, 2014), there is a great diversity of reading comprehension tests, each of which measures certain aspects of this construct better than others (Keenan, Betjemann, & Olson, 2008). In addition to the complexity of this construct, a set of characteristics differentiate this large number of tests, such as time available to answer them, age of the target population, response format (choice or constructed response), purpose (research, teacher assessment, screening or clinical diagnosis), etc. Magliano, Millis, Ozuru and McNamara (2007) state that the decision processes involved in answering choice questions introduce cognitive processes irrelevant to the ability to process written texts in everyday contexts. This type of test introduces into the measurement skills that are not directly related to reading comprehension (Cerdan, Vidal, Martinez, Gilabert, & Gil, 2009; Farr, Pritchard, & Smitten, 1990; Fletcher, 2006; Rupp, Ferne, & Choi, 2006; Ozuru, Rowe, O'Reilly, & McNamara, 2008). However, it should be noted that there is no infallible method for measuring reading comprehension, since each one allows measuring partial aspects of comprehension (Perez, 2005), so it will always be necessary to choose the technique that best fits the particular constraints of the case (Alderson, 2000).

It is of vital importance to identify specifically which cognitive processes a test measures because its usefulness in a given context depends on it. It is from these explanations that a more aesthetic method is used to evaluate reading comprehension in the English Diagnostic Test applied by the University of Costa Rica to first year students. The aim is not only to teach reading, but rather to teach students to enjoy reading and to understand that, just as any language evolves, reading comprehension should also be considered a non-static entity.

Methods

This research adheres to the positivist philosophy. Epistemologically, positivism considers that only one reality is true, and data are generated by statistical methods derived from descriptive or inferential analysis (Rehman & Alharthi, 2016). Therefore, the validity and reliability of instruments are observable, providing evidence of quality of instruments for measuring constructs. This section provides a view of the elaboration of instruments, data collection process, and statistical methods used for the analyses.

Instruments

This study analyzes four instruments for diagnosing the reading comprehension skill. Each instrument consisted of 55 items. Items for each instrument were generated based on the construct's definition and an academic text. Texts were selected in consonance with the following criteria B2+ or C1 level according to the bands of the Common European Framework of Reference for languages (from 1600 up to 1800 words) (Evaluacion en Lenguas Extranjeras, 2022). Instruments were based on different academic texts, so a total of four academic texts were used to develop the items of each instrument. Language specialists at the School of Modern Languages, University of Costa Rica,

created each set of 55 items. Then, another team of specialists reviewed the instruments before being piloted. Instruments were assembled to the virtual platform of the Evaluacion en Lenguas Extranjeras program.

Data collection

Data was collected during March of 2022. Instruments were applied through virtual mode to first-year students at the University of Costa Rica, using the virtual platform of the Evaluacion en Lenguas Extranjeras program and Zoom platform. 1732 first-year students were assessed in this test, randomly assigned to the four instruments.

Regarding the test application, students did the corresponding registration for the tests. Then, they received an email and were invited to a Zoom session. During the application of the test, administrators strictly verify the identity of the students, and they read the instructions of the application. Students were asked to keep their cameras on while the use of any electronic device was banned. Administrators kept rigid observation during the application. Students had two hours to complete the test.

Data Analysis

This research uses two main statistical techniques. Instruments were analyzed separately for anchor items did not exist. To collect evidence of internal validity, this study applied the Exploratory (EFA) and Confirmatory (CFA) Factor Analyses. For the exploratory scenario, tetrachoric correlations were used for estimations, based on results from Freiberg's, Stover, De la Iglesia, & Fernandez Liporace (2013) study. A unidimensional structure fitted the confirmatory analyses.

Evidence of reliability was estimated using two parallel methods. On the one hand, the *Split-Half* method developed by Spearman-Brown. In this case, the set of 50 items were divided into 10 000 random splits to mind the maximum value, and the reported value in this study refers to the second quantile in the distribution of the estimated split-half reliabilities (Revelle, 2022). On the other hand, *Cronbach's alpha* was estimated. Cronbach (1951) established a formula to estimate the lower bound of a test's reliability. Equation 1 shows the formula for estimation.

$$\alpha = \frac{n}{n-1} \left(1 - \frac{\sum_{i=1}^n V_i}{V_w} \right) \quad (1)$$

n is the number of total items in the test, V_i is the variance of item i , and V_w is the variance of the total scores.

For the statistical analyses, test results were analyzed using R in its version 4.1.3 (R Core Team, 2022), using the following packages: nFactors version 2.4.1 (Gilles & Magis, 2020) for EFA and CFA, psych version 2.2.3 (Revelle, 2022) and ShinyItemAnalysis version 1.4.0 (Martinkova & Drabinova, 2018) for CTT estimations.

Results

Table 1 summarizes the results obtained by participants in this study. Means among participants range from 57.1 to 60.2. All instruments have similar standard deviations, ranging from 13.1 to 14.9. Percentages of approval ranges from 20.8 to 28.6, being the third form the one with the lowest percentage of approved examinees (20.8%).

Table 1: Descriptive Statistics of Tests Takers by Form, EDI2022 Test

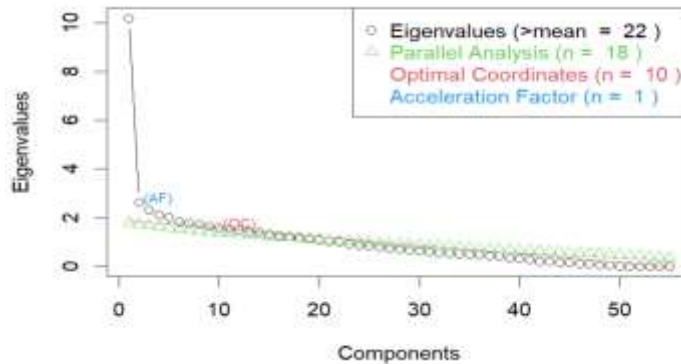
Statistic	Form			
	Set1	Set2	Set3	Set4
Number of Examinees	335	809	413	175
Number of Items	55	55	55	55
<i>Grades Statistics</i>				
Minimum	23.6	29.1	23.6	23.6
I Quartile	49.1	47.3	47.3	48.2
Median	60.0	60.0	58.2	60.0
Mean	58.9	59.8	57.1	60.2
II Quartile	69.1	70.9	69.1	70.9
Maximum	90.9	92.7	85.5	94.6
Standard Deviation	13.1	14.5	13.8	14.9
<i>Percentage of Approval</i>				
Pass	22.1	28.4	20.8	28.6
Fail	77.9	71.6	79.2	71.4

Note. Grades on a scale from 0 to 100.

Source: English Diagnostic Test, School of Modern Languages, University of Costa Rica.

Evidence from the exploratory and confirmatory analyses supports the unidimensionality. Figure 1 shows the number of factors to select after the Exploratory Factor Analysis. The graph discloses one prominent factor over the rest while this is similarly marked by the acceleration factor. This first factor explains 18.52% of the total variance, whereas the rest of the factors represent less than 5% each. Similar structures are found in the other three instruments (see Figure A, Figure B and Figure C in the appendices). Results denote only one noted dimension for measurement in each of the forms.

Figure 1: Scree plot for Eigenvalues after EAF. Form I, EDI 2022



Note: Estimation by tetrachoric correlations.

Source: English Diagnostic Test, School of Modern Languages, University of Costa Rica.

Results from the EFA added to the CFA ones (see Table A in the appendices) demonstrate that each instrument is unidimensional. Accordingly, instruments measured one dimension, allowing the interpretation of the next results.

In reference to the internal consistency, Table 2 shows the main results of the two methods estimated for each of the forms. Three out of the four forms show Cronbach's alpha over 0.8 while only the first form has an alpha of 0.795. Similarly, by means of the Split-Half method, all the forms

show an internal consistency over 0.81. These results demonstrate how the consistency of the instruments used in 2022 aligned with the requirements for high-stake testing.

Table 2: *Indices of Internal Consistency by Form EDI Test 2022*

Statistic	Form			
	I	II	III	IV
Number of examinees	335	809	413	175
<i>Cronbach's Alpha</i>				
Number of Items	55	55	55	55
Initial Cronbach's α	0.795	0.836	0.814	0.845
Excluded Items ^a	9	3	7	4
Final Items	46	52	48	51
Final Cronbach's α	0.832	0.846	0.852	0.867
<i>Split-Half Method</i>				
Lower bound	0.78	0.81	0.82	0.84
Alfa	0.812	0.837	0.846	0.870
Upper bound	0.84	0.86	0.87	0.90

Note. ^a Excluded items based on negative correlation or correlation < 0.5

Source: English Diagnostic Test, English Diagnostic Test, School of Modern Languages, University of Costa Rica.

Discussion

This paper has explored the evidence of validity and reliability in the English Diagnostic Tests at the University of Costa Rica in 2022. This study shows evidence of construct validity by the Exploratory and Confirmatory Factor analyses. The results of the analyses confirmed the existence of unidimensionality. As discussed in Martinez, Hernandez, & Hernandez (2014), tests must respond to the specifications established from their theoretical elaboration. The four forms of EDI test showed to be unidimensional, granting that the instruments measured only one skill, reading comprehension.

The verification of unidimensional for the instruments does not only provide evidence of validity for the EDI Test, but it also propels the interpretation of the instruments' internal consistency. As seen above, most of the instruments exhibited high consistency in their measurements. Especially for tests with high consequences, coefficients of internal consistency over 0.8 become a requirement to guarantee a trustworthy test (ALTE, 2011). This demonstrates that the EDI test is consistent for measuring the reading comprehension of first-year students at the University of Costa Rica.

As seen in this paper, the EDI Test employed virtually four instruments in 2022 with the aim of measuring reading comprehension. These four instruments were elaborated using four different academic readings, and besides the complexity for measuring the reading comprehension construct (Keenan, Betjemann, & Olson, 2008; O'Reilly, Weeks, Sabatini, Halderman, & Steinberg, 2014), the four instruments proved their unidimensionality, as proposed by the Programa de Evaluacion en Lenguas Extranjeras (2022) in its construct definition. Likewise, the EDI's instruments showed high levels of internal consistency. These information evidence high standards of validity and reliability of the test. The program of Evaluacion en Lenguas Extranjeras keeps guaranteeing test takers the development of objective and consistent instruments.

The English Diagnostic Test exemplifies the opportunity to assess reading comprehension skills using valid and consistent tests in virtual environments. In 2022 test administration, the EDI test assessed 1732 first-year students with a sound test. Results from this study have shown the capacity of virtually testing reading comprehension skills, keeping their validity and reliability of the instruments. Future studies should delve about the advantages and disadvantages of virtual test administrations.

The English Diagnostic Test represents an opportunity for many students to advance in the curriculum. For the School of Modern Languages, this test decreases the unsatisfactory enrollment lists for the LM-1030 Reading Strategies in English I course. This has a direct influence in response to the student needs along with the budget ones. Equally important, the EDI test demonstrated to be a valid and reliable test for first-year learners to measure their reading comprehension skills, necessary for the ahead stage of life which is college.

References

- Alexander, P. A. & Fox, E. (2013). A Historical Perspective on Reading Research and Practice, Redux. En D. E. <https://doi.org/10.1598/0710.01>
- Alderson, J. C. (2000). *Assessing Reading*. Reino Unido: Cambridge University Press. <https://doi.org/10.1017/cb09780511732935>
- Alvermann, D. E., Unrau, N. J., & Ruddell, R. B. (2013). *Teoretical Models and Processes of Reading*. Estados Unidos: International Reading Association.
- Anderson, N. (1999). *Exploring Second Language Reading*. Canada: Heinle and Heinle.
- Araya, W. (2017). Conceptualizacion Teorica del Examen de Diagnostico en Ingles de la Escuela de Lenguas Modernas para estudiantes de primer ingreso a la UCR y sus resultados para los anos 2015, 2016 y 2017. *Revista de la Escuela de Lenguas Modernas*, (28). <https://doi.org/10.15517/rlm.vo128.34852>
- Araya, W. (2021). Dominio linguistico en ingles en estudiantes de secundaria para el ano 2019 en Costa Rica. *Revista de Lenguas Modernas*, (34), 1-21. <https://doi.org/10.15517/RLM.Vol34.43364>
- Association of Language Testers in Europe (ALTE). (2011). *Manual for Language Test Development and Examining*. Council of Europe.
- Bachman, L. F. (2007). What is the construct? The dialectic of abilities and contexts in defining constructs in language assessment. In: J. Fox, M. Wesche, D. Bayliss, L. Cheng, C.E. Turner, & C. Doe (Eds.), *Language Testing Reconsidered* (pp. 41-71). Canada: University of Ottawa Press. <https://doi.org/10.2307/j.ctt1ckpccf.9>
- Brizuela Rodriguez, A., Perez Rojas, N., & Rojas Rojas, G. (2020). Validacion de una prueba de comprension de lectura para estudiantes universitarios. *Revista Educacion*, 44(1), 1-14. <https://doi.org/10.15517/revedu.v44i1.34983>
- Cerdán, R., Vidal-Abarca, E., Martinez, T., Gilabert, R., & Gil, L. (2009). Impact of question-answering tasks on search processes and reading comprehension. *Learning and Instruction*, 19(1), 13-27.
- Cerdas-Nunez, D. & Montero-Rojas, E. (2017). Uso del modelo de Rasch para la construccion de tablas de especificaciones: Propuesta metodologica aplicada a una prueba de seleccion universitaria. *Actualidades Investigativas en Educacion*, 17(1), 79-95. <https://doi.org/10.15517/aie.v17i1.27299>
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334.
- Davis, F. B. (1944). Fundamental factors of comprehension in reading. *Psychometrika*, 9(3), 185-197.
- Escuela de Lenguas Modernas, Universidad de Costa Rica (2016). Programa de Curso LM - 1030 Estrategias de Lectura en Ingles I.

- Escuela de Lenguas Modernas, Universidad de Costa Rica (2016). Informe del Examen de Diagnostico en Ingles, 2016.
- Escuela de Lenguas Modernas, Universidad de Costa Rica (2015). Informe del Examen de Diagnostico en Ingles 2015.
- Farr, R., Pritchard, R., & Smitten, B. (1990). A description of what happens when an examinee takes a multiple-choice reading comprehension test. *Journal of Educational Measurement*, 27(3), 209-226.
- Fletcher, J. M. (2006). Measuring reading comprehension. *Scientific Studies of Reading*, 10(3), 323-330. https://doi.org/10.1207/s1532799xssr1003_7
- Freiberg, A., Stover, J. B., De la Iglesia, G., & Fernandez Liporace, M. (2013). Correlaciones policoricas y tetracoricas en estudios factoriales exploratorios y confirmatorios. *Ciencias psicologicas*, 7(2), 151-164. <https://doi.org/10.22235/cp.v7i1.1057>
- Gilles, R. & Magis, D. (2020). nFactors: Parallel analysis and other non graphical solutions to the Cattell Scree Test. <https://CRAN.R-project.org/package=nFactors>
- Green, R. (2019). Item analysis in language assessment. *Quantitative Data Analysis for Language Assessment Volume I: Fundamental Techniques*, 15-29. <https://doi.org/10.4324/9781315187815-2>
- Keenan, J. M., Betjemann, R. S., & Olson, R. K. (2008). Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific Studies of Reading*, 12(3), 281-300. <https://doi.org/10.1080/10888430802132279>
- Kendeou, P., van den Broek, P., Helder, A., & Karlsson, J. (2014). A cognitive view of reading comprehension: implications for reading difficulties. *Learning Disabilities Research & Practice*, 29(1), 10-16. <https://doi.org/10.1111/ldrp.12025>
- Kintsch, W. (2013). Revisiting the construction-integration model of text comprehension and its implications for instruction. In: D.E. Alvermann, N.J. Unrau, & R.B. Ruddell (Eds.), *Theoretical Models and Processes of Reading*. 807-839. International Reading Association.
- Krashen, S. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. *The Modern Language Journal*, 73 (4), 440-464. <https://doi.org/10.1111/j.1540-4781.1989.tb05325.x>
- Landi, N. (2010). An examination of the relationship between reading comprehension, higher level and lower-level reading sub-skills in adults. *Reading and writing*, 23(6), 701-717. <https://doi.org/10.1007/s11145-009-9180-z>
- Leighton, J. P. & Gierl, M. J. (2011). *The learning sciences in educational assessment: The role of cognitive models*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511996276>
- Leslie, L. & Caldwell, J. (2009). Formal and informal measures of reading comprehension. In: S.E. Israel & G.G. Duffy (Eds.), *Handbook of Research on Reading Comprehension*. 403-427. Routledge.
- Llorens, A. C., Gil, L., Abarca, E.V., Martinez, T., Mana, A., & Gilabert, R. (2011). Prueba de competencia lectora para educacion secundaria (CompLEC). *Psichotema*, 23 (4), 808-817.
- Lopez, M. (2010). Estrategias cognitivas del aprendizaje y comprension lectora en estudiantes de quinto grado de secundaria de una institucion educativa del Callao [Tesis de maestria]. Universidad de Costa Rica, San Jose. Recuperado de <https://bit.ly/3orlmdX>
- Mair, P. (2018). *Modern psychometrics with R*. Cham: Springer International Publishing. <https://doi.org/10.1007/978-3-319-93177-7>
- Magliano, J. P., Millis, K. K., Ozuru, Y., & McNamara, D. S. (2007). A multidimensional framework to evaluate reading assessment tools. In: D. McNamara (Ed.), *Reading Comprehension Strategies: Theories, Interventions, and Technologies* 107-136. Lawrence Erlbaum Associates.
- Makuc, M. (2008). Teorias implícitas de los profesores acerca de la comprensión de textos. *Signos*, 41 (68), 403-422. <https://doi.org/10.4067/s0718-09342008000300003>
- Martinez, M. R., Hernandez, M. J., & Hernandez, M. V. (2014). *Psicometría*. Venezuela: Universidad Central de Venezuela.

- Martinkova, P. & Drabinova, A. (2018). ShinyItemAnalysis for teaching psychometrics and to enforce routine analysis of educational tests. *The R Journal*, 10(2), 503-515. <https://doi.org/10.32614/RJ-2018-074>
- McNamara, D. & Magliano, J. (2009). Towards a comprehensive model of comprehension. In: Ross, B. (Ed.), *The psychology of learning and motivation* 297-384. Academic Press. [https://doi.org/10.1016/S0079-7421\(09\)51009-2](https://doi.org/10.1016/S0079-7421(09)51009-2)
- Murillo-Palacios, I. (2008). La teoria de los esquemata y su influencia en la comprension de lectura Universidad de Kansas.
- O'Reilly, T., Weeks, J., Sabatini, J., Halderman, L., & Steinberg, J. (2014). Designing reading comprehension assessments for reading interventions: how a theoretically motivated assessment can serve as an outcome measure. *Educational Psychology Review*, 26(3), 403-424. <https://doi.org/10.1007/s10648-014-9269-z>
- Ozuru, Y., Rowe, M., O'Reilly, T., & McNamara, D. S. (2008). Where's the difficulty in standardized reading tests: The passage or the question?. *Behavior Research Methods*, 40(4), 1001-1015.
- Perez, M. J. (2005). Evaluacion de la comprension lectora: Dificultades y limitaciones. *Revista de Educacion*, 1, 121-138.
- Perfetti, C. & Adlof, S.M. (2012). Reading comprehension: A conceptual framework from word meaning to text meaning. In: J.P. Sabatini, E.R. Albro, & T. O'Reilly. *Measuring Up: Advances in How We Asses Reading Ability*. 3-20. Rowman & Littlefield Education.
- Programa de Evaluacion en Lenguas Extranjeras. (2022). Tabla de especificaciones y constructo prueba de diagnostico de Ingles.
- R Core Team. (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Austria: Vienna. <https://www.R-project.org/>
- Rehman, A. A., & Alharthi, K. (2016). An introduction to research paradigms. *International Journal of Educational Investigations*, 3(8), 51-59.
- Reyes, S., Castillo, A., Zuniga, A., & Llarena, R. (2009). La evaluacion objetiva de la habilidad lectora en la educacion media superior. CENEVAL. <https://bit.ly/2RHodZw>
- Revelle, W. (2022). Procedures for psychological, psychometric, and personality research. <https://CRAN.R-project.org/package=psych>
- Rosenblatt, L. (1977). *The reader, the text, the poem*. Southern Illinois: University Press.
- Rupp, A. A., Ferne, T., & Choi, H. (2006). How assessing reading comprehension with multiple-choice questions shapes the construct: A cognitive processing perspective. *Language testing*, 23(4), 441-474.
- Sanchez Salazar, M. (2002). La lectoescritura en el proceso de aprendizaje de una segunda lengua. *Repertorio Americano*, 13, 292-302.
- Solorzano-Salas, J., & Montero-Rojas, E. (2011). Construccion y validacion de una prueba de comprension de lectura mediante el Modelo de Rasch. *Revista Electronica "Actualidades Investigativas en Educacion"*, 11(2),1-27. <https://doi.org/10.15517/aie.v11i2.10188>
- Spearman, C. (1910). Correlation calculated from faulty data. *British Journal of Psychology*, 3(3), 271. <https://doi.org/10.1111/j.2044-8295.1910.tb00206.x>
- Svetina, D., Gorin, J. S., & Tatsuoka, K. K. (2011). Defining and Comparing the Reading Comprehension Construct: A Cognitive-Psychometric Modeling Approach. *International Journal of Testing*, 11(1), 1-23. <https://doi.org/10.1080/15305058.2010.518261>
- Van den Broek, P. (2012). Individual and Developmental Differences in Reading Comprehension: Assessing Cognitive Processes and Outcomes. In: J. P. Sabatini, E. R. Albro, & T. O'Reilly, *Measuring Up: Advances in How We Asses Reading Ability*, 39-57. Rowman & Littlefield Education.
- Universidad de Costa Rica (2017). UCR en cifras. Consultado el 1 de junio de la pagina web <http://www.ucr.ac.cr/acerca-u/ucr-en-cifras.html>
- Zamora-Araya, J. A. (2015). Analisis de la confiabilidad de los resultados de la prueba de diagnostico matematica en la Universidad Nacional de Costa Rica utilizando el modelo de Rasch. *Actualidades en Psicologia*, 29(119), 153-165. <https://doi.org/10.1177/0033688213500582>

Appendices

Table A
Indices for the Fitted Model, Confirmatory Factor Analysis. EDI Test 2022

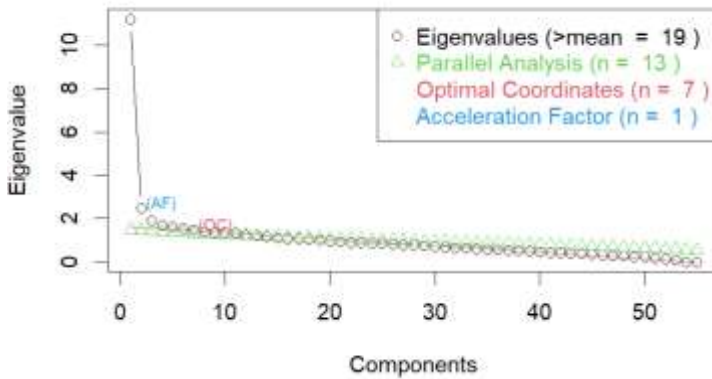
Statistic	Form			
	I	II	III	IV
Total of Observations	335	809	413	175
Kaiser-Meyer-Olkin (KMO)	0.75	0.88	0.81	0.70
<i>Based Model</i>				
Number of Items	55	55	55	55
Chi-Squared Statistic	1645.29	1989.62	1797.37	1915.68
Degrees of Freedom	1377	1430	1430	1430
X^2/df Statistic	1.19	1.39	1.26	1.34
Comparative Fit Index (CFI)	0.822	0.870	0.839	0.665
Tucker Lewis Index (TLI)	0.815	0.865	0.833	0.652
RMSEA	0.024	0.022	0.025	0.044
SRMR	0.051	0.036	0.046	0.071
<i>Adjusted Model</i>				
Number of Items	46	49	46	34
Chi-Squared Statistic	1023.37	1488.80	1052.77	696.54
Degrees of Freedom	945	1127	902	594
X^2/df Statistic	1.08	1.32	1.17	1.17
Comparative Fit Index (CFI)	0.935	0.909	0.916	0.880
Tucker Lewis Index (TLI)	0.932	0.905	0.912	0.873
RMSEA	0.016	0.020	0.020	0.031
SRMR	0.047	0.034	0.044	0.062

Note. KMO index based on tetrachoric correlations.

Source: English Diagnostic Test, English Diagnostic Test, School of Modern Languages, University of Costa Rica.

Figure A

Scree plot for Eigenvalues after EAF. Form II, EDI Test 2022

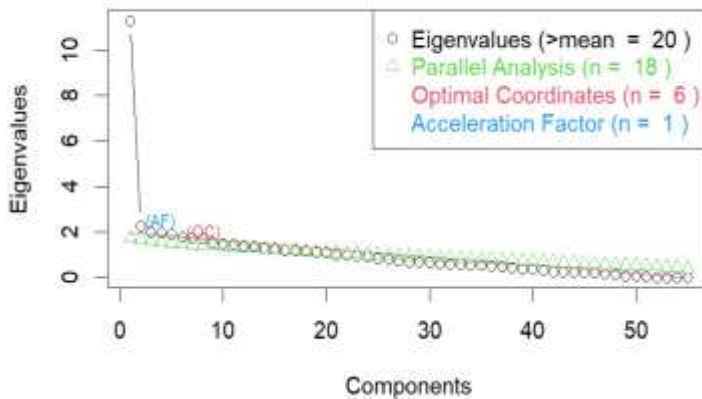


Note: Estimation by tetrachoric correlations.

Source: English Diagnostic Test, School of Modern Languages, University of Costa Rica.

Figure B

Scree plot for Eigenvalues after EAF. Form III, EDI Test 2022

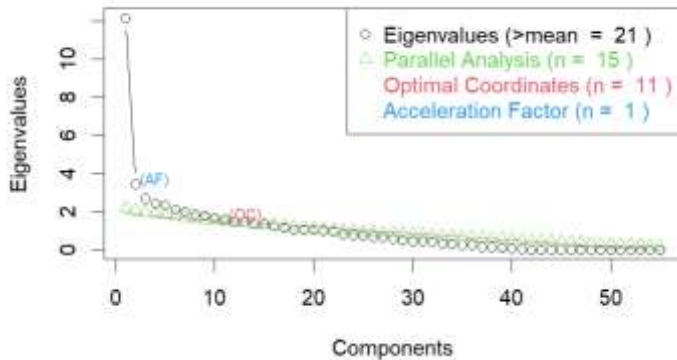


Note: Estimation by tetrachoric correlations.

Source: English Diagnostic Test, School of Modern Languages, University of Costa Rica.

Figure C

Scree plot for Eigenvalues after EAF. Form IV, EDI Test 2022



Note: Estimation by tetrachoric correlations.

Source: English Diagnostic Test, School of Modern Languages, University of Costa Rica.

Biographical notes:

Walter Araya Garita is an English professor and researcher at the University of Costa Rica. He holds a MA in Teaching English as a Foreign Language and a M.Sc. in Administration both from the University of Costa Rica, where he graduated with honors. He also has a specialty in Educational Planning and Administration from the National University of Educational Planning, India. He is the director of the Language Assessment Program at UCR and the secretary of the Latin American Association of Language Testing and Assessment.

José Alejandro Fallas Godínez holds BA in Statistics and a MA in Applied Linguistics with a focus on English for Specific Purposes. He is interested in quantitative research in the measurement field. Alejandro has done research on standardized language testing to collect evidence of the validity and reliability for language tests. He is currently enrolled in a Ph.D Program in Language Assessment at the Laval University.