

Design and qualitative validation of a critical reading test for tertiary education in Colombia

Yeny Alexandra Pulido Aguirre [ORCID Id 0000-0002-2246-8208] y Leonardo Andrés Ferreira [ORCID Id 0009-0001-6469-7199]

Universidad El Bosque, Bogotá. Carrera 9 # 131A- 02, Colombia

ypulidoa@unbosque.edu.co and ferreiraleonardo@unbosque.edu.co

Abstract. This text presents advances of the instrumental research that proposes the creation of a critical reading test aimed at undergraduate students in Colombia. The test was design from the ICFES' Critical Reading test frame of reference and methodology (Item Response Theory, IRT). A bank of items was designed and submitted to an evaluation in search of evidence of its content validity, achieving the approval of 87.5% of the items by the five judges convened (language experts). It was due to the favoring of discussion and attention to the judges' observations. The project will continue with the evaluation of quantitative validity through a pilot test to achieve a valid and reliable test, as well as comparable to the ICFES one.

Keywords: Tests construction, test validity, critical reading, higher education.

1 Introduction

1.1 Problem

Young Colombians perform below average in the PISA reading test according to measurements between 2006 to 2018. At the same time, young persons have a low performance in the Critical Reading test of Saber 11 according to the results between 2017 to 2020 published by Icfes. The situation worsens for those who are culminating their university education, about 60% are in levels 1 and 2 (4 is the highest) of the Critical Reading test (in Saber Pro) according to the 2018 to 2020 measurements from Icfes. This shows that young people entering higher education have a low development of their reading competence and they do not improve it at university.

1.2 Aim

To design a critical reading test for undergraduate students in Colombia.

1.3 Justification

1. Icfes suggests that universities should carry out internal measurements on the development of critical reading skills, among others, to complement its results to monitor the evolution of these skills in their students.

2. A valid critical reading test let to obtain reliable results on students' reading skills and let to identify the skills that require pedagogical intervention.

2 Theoretical Framework

2.1 Critical Reading

Icfes understands critical reading as an indispensable competence for the academic, work, and civic performance of every citizen, regardless of their occupation, profession, and trade". This definition framed in the orientation of Rychen and Salganik (2003, p. 5, cited in Icfes, 2017, pp. 13-14), who from the social level think that critical reading contributes to a successful life and a properly functioning society.

Besides, critical reading consists of understanding the purpose of reading; selecting, understanding, using, and evaluating diverse sources of information; identifying, using con-textualizing and corroborating sources; writing from multiple perspectives (Hyndt and Stahl, 1998, cited in Icfes, 2018, p. 14). It implies the reader has a clear criterion about what they perceive from different sources.

2.2 What is a test?

A test "is a psychological procedure of objective and standardized measurement of a sample of behaviors" (Anastasi, 1998, cited in Meneses et al., 2013, p. 43). According to its purpose, an instrument seeks to diagnose (the current conditions of the evaluated) and predict (relate the current measure to the behavior of people in future situations) (Meneses, et al., 2013, p. 43).

3 Methodos

3.1 Type of study

This is an instrumental research according to Ato, López and Benavente (2013), in which qualitative and quantitative validation will be sought from the psychometric methodology of Item Response Theory (IRT).

3.2 Population and sample

The test has been designed for university students in Colombia and will be piloted with 500 students from a university in Bogota.

3.3 Design

I. Development and evaluation of the test specifications.

II. Development, testing and evaluation of the items.

1. Construction of an item bank.
2. Design of item evaluation forms.
3. Submission of the items to evaluation in search of evidence of content validity.
4. Calculate the concordance between five judges.
5. Elaboration of the conclusions of the trial
6. Assembling and evaluating new evidence forms.
7. Creation of documents to support the administration of the evidence.
8. Application of the test to the sample.
9. Quantitative examination: Consists of performing the unidimensionality test, choosing the model to be used from among one, two or three parameters, estimating the value of these parameters and, finally, checking that the estimated model fits the empirical data (Muñiz, 2018, p. 190).

4 Results and discussion

The project has made progress in:

I. Development and evaluation of the test specifications: The theoretical framework was elaborated based on the reference framework of the Icfes' Critical Reading test, which considers that this competency is composed of three sub-competencies: The first one accounts for the microstructure and the second for the macrostructure, according to Kintsch and van Dijk's conception. The third is the critical one, which consists of reflection and evaluation based on the content of a text. The latter required a conceptual reconstruction because Icfes has not made public what it understands by this. Thus, the list of evidence and the released items related to this sub-competency were used as a basis to identify the theoretical issues that needed to be explained.

Maintaining the Icfes construct ensured the validity of the test in terms of its construct, which was evident in the items created, which were not objected to by the judges in the item evaluation phase.

II. Item development, testing and evaluation:

1. Item bank construction: 64 items were constructed from 5 continuous and discontinuous texts.
2. Design of templates: the expert judgment template of Escobar Pérez and Cuervo-Martínez (2008) was adjusted and used.
3. Submission of the items to evaluation in search of evidence of content validity: five experts in reading and writing with knowledge of the elaboration of critical reading evaluation items, according to the Icfes construct, evaluated each item under the criteria of coherence, relevance, clarity, and sufficiency, giving them a minimum value of 1 and a maximum of 4.

Three evaluation sessions totaling eight hours were held, in which the judges shared their assessment on the data collection matrix and discussed it. This allowed simple adjustments to be made to some items during the sessions and recommendations for the adjustment of others.

In the first session, 30 items were approved, 1 was eliminated and the remaining items were recommended for adjustment. In the second session, 26 items were approved. Thus, of the 64 items, 56 (87.5%) were approved.

It should be noted that, although most of the items passed the evaluation of the majority, the attention to their observations allowed the improvement of many of them. Thus, the judges felt that their observations were heeded while the others found their point of view reasonable and approved of it. This was naturally achieved in an atmosphere of discussion favored by the professionalism of the judges.

5 Conclusions

- Starting from a frame of reference of an existing test accelerates the process of conceptualizing the new test and avoids the search for experts for its assessment.
- The design of items is a long process and depends on a good selection of texts.
- Having judges who do not know each other contributes to a transparent evaluation.
- Fit items paying attention to each judge's comments let to get a great number of approved items.

6 Limitations

Instrument (the test created): it is limited to the conception of critical reading present in the theoretical framework and to the use of the Colombian variety of Spanish.

Risk of bias: Pedrosa, Suárez-Álvarez and García-Cueto (2013, p. 7) warn that the fact that the researcher directs the item evaluation sessions may lead to bias in the judges' assessment; however, efforts were made to avoid this. In any case, it is necessary to wait for the results of the application of the pilot test and the analysis of its results to identify the behavior of the items to have more information.

References

- American Educational Research Association, American Psychological Association, National Council on Measurement in Education. (2014). *Standards for Educational and Psychological Testing of the American Educational Research Association*. Washington: American Educational Research Association.
- Escobar-Pérez, Jazmine; Cuervo-Martínez, Ángela. (2008). Validez de contenido y juicio de expertos: una aproximación a su utilización. *Avances en Medición*, 06, 27-36.

- ICFES. (2018). *Marco de referencia para la evaluación, Icfes: Módulo de lectura crítica Saber 11°, Saber TyT y Saber Pro*. Bogotá: ICFES.
- ICFES. (abril de 2017). *Guía de orientación Saber 11°*. Bogotá: Instituto Colombiano para la Evaluación de la Educación.
- ICFES. (mayo de 2018). *Guía de diseño, producción y aplicación y calificación del examen Saber 11*. Bogotá: ICFES.
- Meneses, A., Barrios, M., Bonilla, A., Cosculluela, A., Lozano, L., Turbay, J., & Valero, S. (2013). *Psicometría*. Barcelona: Editorial UOC.
- Muñiz Fernández, J. (2018). *Introducción a la Psicometría*. Ediciones Pirámide.
- Pedrosa, I., Suárez-Álvarez, J., & García-Cueto, E. (2013). Evidencias sobre la Validez de Contenido: Avances Teóricos y Métodos para su Estimación. *Acción Psicológica, 10(2)*.
- Schleicher, A. (2019). *PISA 2018 Insights and Interpretations*. OECD.