OCNOS

Psychometric validation of a reading comprehension and metacognitive awareness test in university students

Aníbal Puente iD
Universidad Complutense de Madrid, Spain
apuentef@ucm.es

Antonio P. Gutiérrez-de-Blume iD
Georgia Southern University, USA
agutierrez@georgiasouthern.edu

Juan Calderón iD
Universidad Andrés Bello, Chile
juan.calderon@unab.cl

Luis Rojas iD
Universidad San Sebastián, Chile
luis.rojasp@uss.cl

Abstract

Reading comprehension and self-regulation skills are essential for academic success in higher education. This study presents the development and psychometric analysis of the University Reading and Comprehension Strategies Assessment (URCSA), a test designed to examine reading comprehension and metacognitive awareness in first-year university students. A total of 507 education students from a Chilean university participated. Item Response Theory (IRT) analyses of unidimensionality, difficulty, and item discrimination were evaluated. The results show sufficient evidence of internal validity, reliability, and sensitivity of the instrument to differentiate levels of reading skill. Findings concluded that URCSA is a useful tool for identifying strengths and weaknesses in strategic reading comprehension and metacognition in university settings.

Validación psicométrica de un test de comprensión lectora y conciencia metacognitiva en estudiantes universitarios

Aníbal Puente (iD)
Universidad Complutense de Madrid, Spain
apuentef@ucm.es

Antonio P. Gutiérrez-de-Blume (iD)
Georgia Southern University, USA
agutierrez@georgiasouthern.edu

Juan Calderón (iD)
Universidad Andrés Bello, Chile
juan.calderon@unab.cl

Luis Rojas (iD)
Universidad San Sebastián, Chile
luis.rojasp@uss.cl

Resumen

Las habilidades de comprensión lectora y autorregulación son esenciales para el éxito académico en la educación superior. Este estudio presenta el desarrollo y análisis psicométrico de la prueba denominada Estrategias de Lectura y Comprensión Universitaria (ELCU), un test diseñado para evaluar comprensión lectora y conciencia metacognitiva en estudiantes universitarios de primer año. Participaron 507 estudiantes de carreras de educación en una universidad chilena. A través de análisis basados en la Teoría de Respuesta al Ítem (TRI), se evaluó la unidimensionalidad, la dificultad y la discriminación de los ítems. Los resultados muestran evidencia suficiente de validez interna, fiabilidad y sensibilidad del instrumento para diferenciar niveles de habilidad lectora. Se concluye que ELCU es una herramienta útil para identificar fortalezas y debilidades en la comprensión lectora estratégica en contextos universitarios.

Palabras clave: Comprensión lectora; estrategias de lectura; metacognición; validación de pruebas; Educación Superior.

INTRODUCTION

In recent years, the concept of reading comprehension has evolved toward a more complex and multidimensional view, encompassing both cognitive and metacognitive processes involved in meaning-making during reading (Afflerbach et al., 2015; Cain & Oakhill, 2014). This perspective acknowledges that reading is not only about decoding words or understanding sentences, but also about monitoring comprehension, making inferences, linking new information with prior knowledge, and self-regulating the reading process (Schraw, 2009; Soto et al., 2018). In this context, there has been increasing interest in studying metacognitive awareness as an essential component of deep text comprehension, particularly at higher educational levels.

At the university level, reading plays a central role in the acquisition of disciplinary knowledge, critical participation in academic life, and the development of complex cognitive skills (Carlino, 2009; Castles et al., 2018). However, various studies have pointed out that many university students enter with a limited repertoire of reading strategies and little metacognitive awareness of their own reading performance (Vega et al., 2012; Gutiérrez & Schraw, 2015). This represents a significant barrier to their academic progress, especially in programs that require critical and reflective reading of scientific and argumentative texts (O'Reilly et al., 2014).

In this scenario, it becomes essential to have instruments that allow the identification of strengths and weaknesses in reading comprehension and metacognitive strategies among students entering higher education. Although international initiatives such as PISA exist, these assessments target secondary school students and do not address the specificities of the university context (OECD, 2019). Likewise, the available reading comprehension tests in Spanish for young adults show limitations in terms of ecological validity, alignment with university-level texts, and theoretical grounding (Brizuela-Rodríguez et al., 2019; Calderón-Maureira et al., 2020).

Therefore, the following research question is posed: Is it possible to develop a valid and reliable instrument to assess reading comprehension and metacognitive awareness in university students, using texts representative of the academic domain?

The objective of this study is to construct and psychometrically analyze a reading test designed to evaluate these skills in first-year Chilean university students. The instrument, called *University Reading Strategies and Comprehension* (ELCU, by its Spanish acronym), is composed of academic texts from various discourse genres and tasks that require both literal and strategic comprehension. Using Item Response Theory (IRT) models, the study evaluates the test's internal structure, item difficulty and discrimination, and the reliability of the instrument as an initial assessment tool in higher education contexts.

THEORETICAL FRAMEWORK

The university is the context where critical and reflective reading is practiced in order to grasp the logic of a text and its conditions of production; therefore, fragmented and superficial readings are neither sufficient nor useful (García-Sánchez, 2000). Thus, university students must understand discursive construction, be willing to consult different sources when studying a topic, and perform interpretive operations across texts that involve complementing and contrasting information and/or viewpoints. For this reason, students are expected to read and interpret the controversial dimension of discourses, establishing relationships between the text and the author, the text and other texts, and the text and their prior knowledge. For all these reasons and others not presented here, reading comprehension is essential at the university level. This cognitive activity enables a multitude of learning opportunities related both to acquiring specific domain content and to developing cognitive skills (Carlino, 2009). To support this assumption, learning researchers have produced numerous studies that confirm a widely accepted conclusion: there is a positive correlation between reading comprehension level and academic performance (Bashir & Mattoo, 2012; Cimmiyotti, 2013; Oliveira & Santos, 2006).

The main difficulties individuals encounter in reading comprehension activities are: a) losing references, which reveals a reading focused on language forms but not on the meaning relationships established in semantic continuity; b) difficulty interacting with the text's structure as proposed by the author, resulting in a reading based solely on the reader's own schemas; c) challenges identifying the key ideas that unify the text's information and understanding how the writer connects them through a specific rhetorical structure; d) difficulty understanding situational contexts and the communicative situation that generates the text, which helps identify the author's purposes (to persuade, inform, seduce, etc.); and e) difficulty distancing oneself and self-regulating the comprehension process (Martínez, 1997). Narvaja et al. (2003) suggest that it is very common for university students to read differently from how their professors expect. This often happens because students are not explicitly taught: a) how they should read; b) the intended reading objectives; and c) they are not trained in reading comprehension strategies, as it is assumed they developed them during secondary education. The authors argue that university students tend to reproduce traditional forms of reading. Reading fragments in a decontextualized manner, focusing on identifying the topic without considering the author's intent, is an example of this. Thus, reading tends to center on indiscriminately collecting data, with little attention given to the author's intention, which is expressed through certain linguistic, semantic, or paraverbal features. Identifying the author's intent is linked to the level of text comprehension and the variables that influence it. Therefore, it is essential that the reader not only understand what the author says, but also why.

Reading comprehension

Reading comprehension is not a single process but rather involves many, implying cognitive processes that operate across different types of knowledge. Despite the complexity of this concept, one core idea stands out: comprehension occurs when the reader constructs one or more mental representations of a text (Kintsch & Rawson, 2005) and creates a situational model (Van Dijk & Kintsch, 1983); this model is mainly characterized by its multidimensional and meaningful reference. In terms of reference, authors like Johnson-Laird (1983), Kintsch (1988; 1998), Van Dijk (2006), and Van Dijk and Kintsch (1983) state that if a reader cannot find a clear reference to understand the situation the text refers to, comprehension fails and memory is poor (Herrada-Valverde & Herrada-Valverde, 2017). In short, the situational model results from deep and successful comprehension (Castles et al., 2018). Comprehension and representation processes occur at various levels: word level (lexical processing), sentence level (syntactic processing), and text level (referential and inferential mapping). All these levels interact with the reader's knowledge to produce the text's situational model (Puente, 1991).

Perfetti and Adlof (2012) present the components of the reading process in a more orderly manner than typically occurs. Their description of real-time reading is fairly dynamic and accurate, facilitating the evaluation of reading operations. Thus, two processes are included: 1) word identification and 2) language processing mechanisms for constructing meaning. In this representation, all processes and knowledge sources become focal points for analysis and assessment of comprehension ability. The framework clarifies which components can be evaluated and which cannot. An assessment with clear focal points is helpful for informing stakeholders (e.g., teachers, parents, administrators, and researchers).

Word identification is a critical component of reading comprehension: there are strong, abundant, and important correlations between word identification skills and reading comprehension across all age groups, including adulthood (Adlof et al., 2006; Bell & Perfetti, 1994; Braze et al., 2007; Perfetti, 1985; Sabatini, 2002; 2003). However, while any component, including word identification, may be necessary, it might not be sufficient on its own for comprehension. Some components may not be required for shallow comprehension. Until recently, most research on reading difficulties focused solely on word reading. In recent years, however, it has become clear that some children and adults have specific issues with reading comprehension. That is, they perform poorly in comprehension despite good word recognition skills (Catts et al., 2006; Hart, 2005; Landi, 2010; Nation & Snowling, 1999; Yuill & Oakhill, 1991).

Measuring reading comprehension at the university level

The PISA tests, administered every three years, assess key skills in students, including reading literacy, defined as the ability to understand, use, and reflect on texts to achieve personal goals and actively participate in society (Feito, 2008). This competency involves integrating texts with prior knowledge, assessing their reliability and relevance (OECD, 2019), and includes digital and hypertext reading skills.

In the university context, reading texts is essential for knowledge acquisition. Difficulty understanding social, scientific, or technical documents may limit academic performance. However, there is a shortage of reading comprehension tests in Spanish aimed at university students, and many of the existing ones lack rigorous measurement models. This study presents progress in validating a test for first-year students in Chile.

University texts feature more complex structures than school texts: technical vocabulary, dense grammatical structures, and varied organization, requiring more advanced reading skills (Brizuela-Rodríguez et al., 2019). Unlike children—whose reading comprehension has been widely measured—university students' skills have been less explored (Landi, 2010), and existing tests often focus on lower-level skills.

Language assessment has included both "indirect" strategies (items with incomplete texts) and "direct" strategies (realistic scenarios), with indicators developed from various theoretical frameworks (Bachman, 2007; Leslie & Caldwell, 2009; Perfetti & Adlof, 2012). Reading comprehension has been conceptualized as a skill, activation, or processing (Keenan et al., 2008), leading to a wide diversity of tests.

Different countries have developed tests based on what they consider essential for a reader (Martínez Rizo, 2009). Since Davis (1944), there has been interest in identifying the skills needed to respond correctly to comprehension items. However, not all tests measure the same aspects (Keenan et al., 2008), highlighting the complexity of the construct, which involves multiple cognitive processes (O'Reilly et al., 2014).

Factors such as response time, examinee age, item type, and test purpose (teaching evaluation, research, selection, diagnosis) distinguish different tests. Magliano et al. (2007) warn that multiple-choice questions introduce cognitive processes unrelated to real comprehension, mixing in irrelevant skills (Cerdán et al., 2009; Farr et al., 1990; Fletcher, 2006; Rupp et al., 2006; Ozuru et al., 2008).

There is no single method for measuring reading comprehension. Each test assesses partial aspects of the process (Pérez-Zorrilla, 2005), so its selection should reflect the context and specific processes to be observed (Alderson, 2000). In higher education admissions, tests must measure higher-order skills (Jackson, 2005). Based on this framework, this study was developed to construct an instrument for Spanish-speaking university students.

Moos and Azevedo (2008), in a study with 49 students, analyzed the relationship between prior knowledge, learning self-regulation, and performance. They used think-aloud protocols and categorized processes into planning, strategy use, and monitoring. Students with greater prior knowledge planned and monitored better, scoring higher. In contrast, those with less knowledge used more strategies, mostly simple summaries.

Reading metacomprehension is defined as the reader's ability to reflect on and regulate their comprehension (Schraw, 2009; Soto et al., 2018). It includes processes such as pre-planning, real-time monitoring, and post-reading evaluation (Flavell, 1979; Schraw & Moshman, 1995). Those with high metacomprehension adapt their strategies more effectively and transfer what they've read to new contexts (Vega et al., 2012; Moos & Azevedo, 2008). These findings align with studies highlighting the importance of metacognitive strategies in interpreting complex texts (Puente et al., 2019).

The ESCOLA scale (Jiménez-Rodríguez et al., 2009) evaluates aspects of reading metacomprehension in children aged 8 to 13. It is recommended to complement it with observation and other tests, as it is not meant for diagnosis but rather to identify deficiencies in reading awareness. The test is structured around processes (planning, monitoring, and evaluation) and variables (person, task, and text), supported

by theory. It was developed by an interdisciplinary team with years of work and extensive literature review.

ESCOLA's psychometric studies indicate good reliability, discriminant validity compared to other tests, and convergent validity with MARSI. It has been adapted into Spanish with two parallel versions using Samejima's IRT model, and its properties have been evaluated with samples from Spain and Argentina. The test is available in long (56 items) and short (28 items) versions, adjusted by age. It is easy to apply and interpret, and includes supporting materials for its use.

Vega et al. (2012) studied reading self-regulation in university students reading multiple texts on bacterial resistance. They used think-aloud techniques to gather cognitive processes and assessed comprehension. They found little planning, poor monitoring in relation to reading goals, and high use of superficial strategies, even with irrelevant information. In contrast, deep comprehension strategies were used minimally. Ultimately, students performed better on tasks involving superficial comprehension than on those requiring knowledge transfer.

The present study

Based on the reviewed literature, this study aims to validate a diagnostic assessment of reading metacomprehension in university students. The expectation is that the diagnostic test effectively measures students' metacomprehension at more advanced developmental stages.

METHOD

Participants and sampling

Participants were students (N = 507) selected through convenience sampling and enrolled in education degree programs in various specialties (table 1). Most had taken the PSU, a standardized university admission test featuring multiple-choice questions and closed responses, covering mandatory curricular content. It does not include opinion-based tasks, fieldwork, written reports, debates, or presentations. The average age of participants was 19 years and 2 months. All were first-year students, with 70% identifying as female and 30% as male. Interestingly, only in Physical Education were participants evenly distributed across gender identities. PSU scores ranged from 384 to 646 points. Scores below 500 are typically not accepted in most public universities and scientific fields, and are only rarely accepted at some private institutions.

Table 1

Characterization of participants by program of study

Program of Study	N
General Basic Education	16
Music Education	7
Early Childhood Education	53
Physical Education Teaching	94
English Language Teaching	73
Secondary Education Teaching for Degree Holders	233
Psychopedagogy	30
Other Program	1

Program of Study	N
Total	**507**

Instruments and materials

The initial reading comprehension test consists of six texts. Both the texts and the items were previously validated through psychometric testing, which allowed for the evaluation of the instrument's reliability and validity. This task was previously administered to 36 engineering students and 219 education students (Rojas et al., 2024). Regarding the instructions for completing the test, a series of questions must be answered according to the guidelines available at Rojas et al. (2025a).

Remote administration of the ELCU reading test

The test was administered remotely to the participating subjects between March and June 2020. Participants were given access to the SIGECOL platform, where each individual had a personal account that allowed them to complete the ELCU test, which had been previously uploaded to the system. Students were provided with a document containing detailed instructions for completing the test, which can be accessed at Rojas et al. (2025b).

Data analysis

The data were first tested for the necessary statistical assumptions and examined for extreme outliers. The data met all statistical assumptions, and no extreme outliers were identified that could undermine the reliability of the dataset. Descriptive statistics for the items can be found in table 2.

Table 2

Descriptive statistics of the items for the ELCU measure

Item	Min.	Max.	M	Med.	SD	Skewness	Kurtosis	IQR
Text1_Q1	0	2	1.69	2	0.66	-1,92	2.05	0
Text1_Q2	0	2	1.66	2	0.64	-1.7	1.5	0
Text1_Q3	0	2	1.09	1	0.46	0.35	1.54	0
Text1_Q4	0	2	1.35	2	0.84	-0.73	-1.18	1
Text2_Q1	0	2	1.58	2	0.66	-1.32	0.46	1
Text2_Q2	0	2	1.44	2	0.87	-1	-0.92	2
Text2_Q3	0	2	1.65	2	0.52	-1.13	0.24	1
Text2_Q4	0	2	1.85	2	0.47	-0.99	1.06	2
Text4_Q2	0	2	0.89	1	0.48	-0.025	0.92	0
Text4_Q3	0	2	1.01	1	0.83	-0.02	-1.55	2
Text4_Q4	0	2	1.66	2	0.66	-1.7	1.41	0
Text6_Q1	0	2	1.03	1	0.88	-0.06	-1.69	2
Text6_Q2	0	2	1.18	2	0.79	-0.32	-1.34	1
Text6_Q3	0	2	1.18	1	0.54	0.09	-0.01	1
Text7_Q1	0	8	1.51	1	0.65	1.7	1.87	2

Aníbal Puente, Antonio P. Gutiérrez-de-Blume, Juan Calderón, Luis Rojas

Item	Min.	Max.	M	Med.	SD	Skewness	Kurtosis	IQR
Text7_Q2	0	8	1.13	1	0.59	1.33	1.66	2
Text7_Q3	0	8	1.11	1	0.6	0.56	0.72	1
Text7_Q4	0	8	2.22	2	0.82	0.72	0.61	4

N = 507 participating students.

The data were subjected to Item Response Theory (IRT) analysis using version 4.2.2 of jMetrik. jMetrik v4.2.2, a software specialized in psychometric analysis and Rasch modeling, was used due to its capability to perform precise estimations of item difficulty and discrimination, as well as unidimensionality and reliability analyses. These IRT models allow for the evaluation of each individual test item and how well each item provides information about a person's assumed ability in an underlying trait, in this case, reading metacomprehension and metacognitive strategies (Rasch, 1960). According to Rasch,

> "... A person with greater ability than another should have a higher probability of correctly answering any item of the type in question, and similarly, an item that is more difficult than another means that for any person, the probability of answering the second item correctly is lower" (1960, p. 117).

For this purpose, jMetrik produces descriptive statistics for items, a Principal Components Analysis (PCA) of the Standardized Residuals (ZRED) to test unidimensionality, item difficulty estimates and fit statistics (expressed as logits), and item/person separation reliability. The Rasch model was also chosen for its capacity to provide invariant estimates of item difficulty and subject ability, as well as for its suitability in validating the internal structure of unidimensional instruments in educational research (Bond & Fox, 2015).

RESULTS

Tables 3 and 4 present the item correlation matrix for the ELCU test. A general view of the correlation pattern shows that the items are largely orthogonal, as most correlations were weak and non-significant, except for the relationship between Text2_Q1 and Text2_Q2, $r = 0.19$, and the relationships among all items in the metacognitive strategies section of the measure (Text 7), where correlations ranged between $r = 0.25$ and $r = 0.46$.

Table 3

Item correlation matrix for the ELCU measure (Items 1 to 9)

Item	1	2	3	4	5	6	7	8	9
1. Text1_Q1	-	0.15	0.004	-0.02	-0.06	0.02	-0.03	0.06	0.003
2. Text1_Q2		-	-0.03	0.05	0.02	0.06	0.02	0.09	0.12
3. Text1_Q3			-	0.11	0.06	0.05	-0.01	0.01	-0.03
4. Text1_Q4				-	0.04	0.01	0.03	0.08	0.04
5. Text2_Q1					-	.19*	0.03	0.11	0.003
6. Text2_Q2						-	-0.02	0.09	0.1
7. Text2_Q3							-	0.05	0.04

Item	1	2	3	4	5	6	7	8	9
8. Text2_Q4								-	0.11

Table 4

Item correlation matrix for the ELCU measure (Items 10 to 18)

Item	10	11	12	13	14	15	16	17	18
1. Text1_Q1	-0.05	-0.01	0.05	0.05	0.07	-0.01	0.02	0.09	0.02
2. Text1_Q2	0.05	0.02	0.08	0.08	0.03	-0.12	-0.07	-0.06	-0.06
3. Text1_Q3	-0.09	0.04	0.03	-0.004	-0.03	0.06	-0.05	0.02	-0.02
4. Text1_Q4	0.12	0.03	-0.001	0.03	0.02	-0.08	-0.08	-0.04	-0.06
5. Text2_Q1	0.08	0.09	0.04	-0.01	-0.003	-0.001	-0.07	-0.09	-0.06
6. Text2_Q2	0.07	-0.002	0.03	0	-0.08	-0.04	0.01	0.04	0.08
7. Text2_Q3	0.04	-0.06	-0.01	0.03	-0.01	-0.05	-0.04	-0.01	-0.03
8. Text2_Q4	0.04	0.04	0.04	0.003	-0.05	-0.06	-0.16	-0.09	-0.08
9. Text4_Q2	0.01	0.11	0.1	0.02	-0.003	0.02	-0.03	-0.04	0.002
10. Text4_Q3	-	-0.01	0.11	0.02	-0.01	-0.02	0.01	0	0.05
11. Text4_Q4		-	0.03	0.03	0.05	-0.01	-0.05	-0.03	-0.01
12. Text6_Q1			-	0.09	0.01	0	-0.07	-0.04	-0.02
13. Text6_Q2				-	0.13	-0.02	-0.01	0.01	-0.02
14. Text6_Q3					-	-0.04	-0.01	0	0.01
15. Text7_Q1						-	.45**	.41**	.34**
16. Text7_Q2							-	.46**	.25*
17. Text7_Q3								-	.35**
18. Text7_Q4									-

The results of the PCA of standardized residuals showed that the ELCU measure provided evidence of unidimensionality for the six text items and for the metacognitive strategy items that make up the seventh text. The PCA results are presented in table 5. However, there were some items that did not adequately fit the observed data. These misfitting items are discussed below.

Table 5

Principal components analysis results of standardized residuals (ZRED)

Variance Components	Raw Variance	% of Variance
Variance Explained by Measures	15.50	52.60
Variance Explained by Individuals	6.40	21.60
Variance Explained by Items	9.20	31.00
Total Unexplained Variance	14.00	47.40
Total Variance in Observations	29.50	100.00

Variance Components	Raw Variance	% of Variance
Unexplained Variance in the First Contrast	1.90	6.40

N = 507 participating students.

Interestingly, neither the items from Text 3 nor those from Text 5 demonstrated adequate fit to the observed data, as both contributed excessive unexplained variance to the PCA solution—a finding corroborated by the person/item map. The same was true for several other items, including Text1_Q5, Text2_Q5, and Text4_Q1. The total explained variance of the items and the person/item map improved with the removal of these items. Therefore, these items were excluded from further analyses.

The results for item discrimination and individual item difficulty for the remaining well-fitting items are presented in table 6. Item difficulty ranged from a minimum of 1.03 (easiest item, Text1_Q3) to a maximum of 2.22 (most difficult item, Text7_Q4), with the other items falling between these two extremes. These results show that the ELCU measure includes items with an appropriate distribution of item difficulty.

Regarding item discrimination, the results in table 6 show that the items tended to discriminate slightly below average, since the typical item discrimination in the Rasch model is 1 (indicating average discrimination), with values above 1 indicating over-discrimination. Ideally, values close to 1 suggest adequate item discrimination. For the ELCU measure, item discrimination ranged from 0.69 (low) to 1.12 (high), demonstrating a suitable distribution of item discrimination.

Table 6

Rasch analysis of individual items by item statistics

Text by Item	Difficulty	SD	Discrimination
Text 1			
Q1	1.70	0.66	0.73
Q2	1.66	0.56	0.82
Q3	1.08	0.46	0.69
Q4	1.35	0.83	0.71
Text 2			
Q1	1.58	0.66	0.64
Q2	1.45	0.87	0.79
Q3	1.65	0.53	0.88
Q4	1.86	0.48	0.86
Text 4			
Q2	1.89	0.48	0.73
Q3	1.61	0.83	0.78
Q4	1.67	0.65	0.81
Text 6			
Q1	1.33	0.87	0.75
Q2	1.77	0.79	0.73
Q3	1.78	0.54	1.12
Text 7			
Q1	1.51	0.65	0.73
Q2	1.62	0.59	0.90

Text by Item	Difficulty	SD	Discrimination
Q3	1.71	0.60	0.83
Q4	2.22	0.82	0.88

N = 507 participating students.

The test-level statistics for the series of texts are presented in table 7. The item separation indices and person separation reliability are particularly relevant. These range from 0 to 1, and like Cronbach's alpha, indicate higher reliability as values approach 1. Item separation reliability indicates the degree to which item difficulties—here, within the text items—differ from one another. Person separation reliability assesses the extent to which the measure—in this case, the text items—differentiates individuals' assumed abilities in reading metacomprehension or metacognitive strategies.

As highlighted in table 7, the item separation reliability coefficients ranged from 0.75 to 0.90, and the person separation reliability coefficients ranged from 0.79 to 0.89. Therefore, the various texts in the ELCU are not only capable of adequately distinguishing item difficulty but also of differentiating individuals' assumed abilities in reading metacomprehension and metacognitive strategies.

Table 7

Rasch analysis of individual items by test-level statistics

Text	Min.	Max.	M	Med.	SD	Interquartile Range	Skewness	Kurtosis	Item Separation Reliability	Person Separation Reliability
1	1	8	5.80	6	1.42	2	-0.82	0.37	0.75	0.79
2	0	8	6.54	7	1.45	2	-1.09	1.44	0.82	0.86
4	0	6	3.56	4	1.19	1	-0.47	0.22	0.86	0.83
6	0	6	3.38	3	1.39	2	-0.05	-0.68	0.80	0.82
7	0	32	5.97	5	1.85	4	0.21	1.37	0.90	0.89

N = 507 participating students.

In summary, despite the presence of some misfitting items, the ELCU is an adequate measure of reading metacomprehension and metacognitive strategies during reading. With sufficient evidence of unidimensionality for the final items of Texts 1, 2, 4, and 6 regarding reading metacomprehension, and Text 7 regarding metacognitive strategies during reading, an average reliability of r = 0.72, and acceptable item/person separation reliabilities, the final ELCU measure can be used for its intended purpose: to provide reliable and valid information on university students' reading metacomprehension ability and metacognitive strategies during reading.

DISCUSSION

Reading comprehension and metacognitive awareness are fundamental competencies in the university context, especially in training programs where students are expected not only to access text content but also to analyze, interpret, and use it to make informed decisions. Various studies have shown that deep comprehension requires the reader to self-regulate their reading process, identify comprehension difficulties, and activate strategies to overcome them (Cain & Oakhill, 2014; Schraw, 2009). This study aligns with that perspective by presenting the development and psychometric analysis of the ELCU, a test aimed at assessing both reading comprehension and the metacognitive strategies employed during academic reading.

The results of the study show robust evidence of unidimensionality, reliability, and discrimination for most items in the instrument, particularly in Texts 1, 2, 4, 6, and 7. This validation supports the claim that the ELCU is a relevant tool for identifying students' reading profiles upon entering university. Unlike other existing instruments that focus on more basic reading skills or are designed for school-level assessments (Brizuela-Rodríguez et al., 2019), the ELCU targets higher-order processes such as critical interpretation, meaning construction, and strategic use of knowledge, in alignment with contemporary models of academic reading (Castles et al., 2018; Soto et al., 2018).

In addition, the use of the Rasch model enabled a detailed analysis of item behavior, identification of items that did not conform to the expected structure (such as those from Texts 3 and 5), and informed decisions regarding their exclusion. Although widely used internationally (Bond & Fox, 2015), this approach is still rare in reading assessment studies in Latin America, making it a methodological innovation that strengthens the instrument's internal validity.

From a practical standpoint, the ELCU provides higher education institutions with a useful tool for early diagnosis of students' reading strengths and weaknesses, which can help design more effective pedagogical supports. The ability to integrate individual and group results can also offer valuable insights for institutional decision-making in academic support programs, development of transversal competencies, or leveling strategies.

However, the study presents limitations that must be considered. First, the use of convenience sampling limits the generalizability of the results, so future research should replicate the study across different universities, regions, and academic disciplines. Second, although the instrument includes texts of various types and complexities, it is necessary to continue refining those items that did not show adequate statistical fit or generated erratic response patterns. Additionally, the potential differential item functioning by variables such as gender, type of degree, or prior reading history was not analyzed—an interesting direction for future studies.

Looking ahead, it is suggested to: (1) expand the application of the instrument to students from other disciplines to assess its performance in more diverse contexts; (2) incorporate digital technologies to enable adaptive testing, including automated feedback; and (3) conduct longitudinal studies linking ELCU results to students' academic performance throughout their university education. It would also be relevant to explore students' perceptions of the test items and how they relate to their academic reading habits.

In summary, this study offers a significant contribution at theoretical, methodological, and practical levels. Theoretically, it strengthens the understanding of reading metacomprehension as a measurable construct in university populations. Methodologically, it provides a rigorous, replicable, and contextualized validation model. Practically, it delivers a flexible instrument with large-scale implementation potential, useful for supporting inclusion, equity, and the strengthening of advanced reading competencies in higher education.

AUTHOR CONTRIBUTIONS

Aníbal Puente: Project administration; Formal analysis; Conceptualization; Data curation; Writing – original draft; Writing – review & editing; Investigation; Methodology; Resources; Supervision; Validation; Visualization; Funding acquisition.

Antonio P. Gutiérrez-de-Blume: Formal analysis; Data curation; Writing – review & editing; Investigation; Validation; Visualization.

Juan Calderón: Formal analysis; Data curation; Writing – review & editing; Investigation; Software; Validation; Visualization.

Luis Rojas: Formal analysis; Data curation; Writing – review & editing; Investigation; Software; Validation; Visualization.

REFERENCES

Adlof, S. M., Catts, H. W., & Little, T. D (2006). Should the simple view of reading include a fluency component? *Reading and Writing, 19*(9), 933-958. https://doi.org/10.1007/s11145-006-9024-z

Afflerbach, P., Cho, B. Y., & Kim, J. Y. (2015). Conceptualizing and assessing higher-order thinking in reading. *Theory Into Practice, 54*(3), 203-212. https://doi.org/10.1080/00405841.2015.1044367

Alderson, J. C. (2000). *Assessing reading.* United Kingdom: Cambridge University Press. https://doi.org/10.1017/CBO9780511732935

Bachman, M. K. (2007). "Who Cares?:" Novel reading, narrative attachment disorder, and the case of the old curiosity shop. *Journal of Narrative Theory, 37*(2), 296-325. https://doi.org/10.1353/jnt.2008.0003

Bashir, I., & Mattoo, N. H. (2012). A study-on-study habits and academic performance among adolescents 14-19 years. *International Journal of Social Science Tomorrow, 1*(5), 1-5.

Bell, L. C., & Perfetti, C. A. (1994). Reading skill: Some adult comparisons. *Journal of Educational Psychology, 86*(2), 244-255. https://doi.org/10.1037/0022-0663.86.2.244

Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences* (3rd ed.). Routledge. https://doi.org/10.4324/9781315814698

Braze, D., Tabor, W. Shankweiler, D., & Mencl, W. E. (2007). Speaking up for vocabulary reading skill differences in young adults. *Journal of Learning Disabilities, 40*(3), 226-243. https://doi.org/10.1177/00222194070400030401

Brizuela-Rodríguez, A., Pérez-Rojas, N., & Rojas-Rojas, G. (2019). Validación de una prueba de comprensión lectora para estudiantes universitarios. *Revista Educación, 44*(1). https://doi.org/10.15517/revedu.v44i1.34983

Cain, K., & Oakhill, J. (2014). *Children's comprehension problems in oral and written language: A cognitive perspective.* Guilford Press.

Calderón-Maureira, J. F., Puente, A., & Mendoza-Lira, M. (2020). Academic literacy in engineering students: Formulation of an instrument to assess reading comprehension and metacognitive strategies. En *12th International Conference on Education and New Learning Technologies, EduLearn20 Procedings.* https://doi.org/10.21125/edulearn.2020.1900

Carlino, P. (2009). Leer y escribir en la universidad, una nueva cultura. ¿Por qué es necesaria la alfabetización académica? *Página y Signos, 3*(5), 13-52. https://n2t.net/ark:/13683/p1s1/q62

Castles, A., Rastle, K., & Nation, K. (2018). Ending the reading wars: Reading acquisition from novice to expert. *Psychological Science in Public Interest, 19*(1), 5-51. https://doi.org/10.1177/1529100618772271

Catts, H. W., Adlof, S. M., & Weismer, S. E. (2006). Language deficits in poor comprehenders: A case for the simple wiew of reading. *Journal of Speech, Language and Hearing Research, 49*(2). https://doi.org/10.1044/1092-4388(2006/023)

Cerdán, R., Vidal-Abarca, E., Martínez, T., Gilabert, R., & Gil, L. (2009). Impact of question-answering tasks on search processes and reading comprehension. *Learning and Instruction, 19*(1), 13-27. https://doi.org/10.1016/j.learninstruc.2007.12.003

Cimmiyotti, C. B. (2013). Impact of reading ability on academic performance at the primary level. [Master's Thesis]. https://doi.org/10.33015/dominican.edu/2013.edu.18

Davis, F. B. (1944). Fundamental factors of comprehension in reading. *Psychometrika, 9*(3), 185-197.

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. https://doi.org/10.1111/j.1745-3984.1990.tb00744.x

Feito, R. (2008). Competencias educativas: Hacia un aprendizaje genuino. *Andalucía Educativa*, 66, 24-26.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist, 34*(10), 906-911. https://doi.org/10.1037/0003-066X.34.10.906

Fletcher, J. M. (2006). Measuring reading comprehension. *Scientific Studies of Reading, 10*(3), 323-330. https://doi.org/10.1207/s1532799xssr1003_7

García-Sánchez, J. N. (2000). Evaluación e intervención en las funciones verbales, lectura y escritura. En J. N. García-Sánchez (Coord). *De la psicología de la instrucción a las necesidades curriculares* (pp. 189-201). Oikos Tau: Barcelona.

Gutierrez, A. P., & Schraw, G. (2015). Effects of strategy training and incentives on students' performance, confidence, and calibration. *The Journal of Experimental Education*, 83, 386-404. https://doi.org/10.1080/00220973.2014.907230

Hart, L. A. (2005). *A training study using an artificial orthography: Effects of reading experience, lexical quality, and text comprehension in L1 and L2* [Unpublished doctoral dissertation. University of Pittsburgh]. https://d-scholarship.pitt.edu/8235/1/Dissertation%5B1%5D.Hart.04.01.2005c.pdf

Herrada-Valverde, G., & Herrada-Valverde, R. I. (2017). Factores que influyen en la comprensión lectora de hipertexto. *Ocnos, 16*(2), 7-16. https://doi.org/10.18239/ocnos_2017.16.2.1287

Jackson, N. E. (2005). Are university students' component reading skills related to their text comprehension and academic achievement? *Learning and Individual Differences, 15*(2), 113-139. https://doi.org/10.1016/j.lindif.2004.11.001

Jiménez-Rodríguez, V., Puente, A., Alvarado-Izquierdo, J. M., & Arrebillaga-Durante, L. (2009). Medición de estrategias metacognitivas mediante la Escala de Conciencia Lectora: ESCOLA. *Electronic Journal of Research in Educational Psychology, 7*(18), 779-804. https://doi.org/10.25115/ejrep.v7i18.1326

Johnson-Laird, P. N. (1983). *Mental models. Towards a cognitive science of language, inference, and consciousness.* Cambridge University Press.

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

Kintsch, W. (1988). The role of knowledge in discourse comprehension: a construction-integration model. *Psychological Review, 95*(2), 162-182. https://doi.org/10.1037/0033-295X.95.2.163

Kintsch, W. (1998). *Comprehension. A paradigm for cognition.* Cambridge: Cambridge University Press.

Kintsch, W., & Rawson, K. A. (2005). Comprehension. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 209-226). Blackwell Publishing. https://doi.org/10.1002/9780470757642.ch12

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, 701–717. https://doi.org/10.1007/s11145-009-9180-z

Leslie, L., & Caldwell, J. S. (2009). Formal and informal comprehension assessment. In Israel, S., Duffy, G. (Eds.), *Handbook of reading comprehension* (pp. 403–427). Erlbaum.

Magliano, J. P., Millis, K. K., Ozuru, Y. y McNamara, D. S. (2007). A multidimensional framework to evaluate reading assessment tools. En D. S. McNamara (Ed.), *Reading comprehension strategies: theories, interventions, and technologies* (pp. 107-136). Lawrence Erlbaum Associates.

Martínez, M. C. (1997). El desarrollo de estrategias discursivas a nivel universitario. En M.C. Martínez (Comp.). *Los procesos de la lectura y la escritura* (pp. 11-41). Universidad del Valle.

Martínez Rizo, F. (2009). Evaluación formativa en aula y evaluación a gran escala: hacia un sistema más equilibrado. *Revista electrónica de investigación educativa*, *11*(2), 1-18.

Moos, D. C., & Azevedo, R. (2008). Self-regulated learning with hypermedia: The role of prior domain knowledge. *Contemporary Educational Psychology*, 33, 270-298. https://doi.org/10.1016/j.cedpsych.2007.03.001

Narvaja, E., Di-Stefano, M., & Pereira, C. (2003). *La lectura y la escritura en la universidad*. Eudeba.

Nation, K., & Snowling, M. J. (1999). Developmental differences in sensitivity to semantic relations among good and poor comprehenders: Evidence from semantic priming. *Cognition, 70*(1), B1–B13. https://doi.org/10.1016/s0010-0277(99)00004-9

OECD. (2019). *PISA 2018 results (volume I): What students know and can do.*

Oliveira, K. L., & Santos, A. A. A. (2006). Compreensão de textos e desempenho acadêmico. *Psic: Revista de Psicologia da Vetor Editora*, *9*(1), 19-27.

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. https://doi.org/10.3758/BRM.40.4.1001

Pérez-Zorrilla, M. J. (2005). Evaluación de la comprensión lectora: Dificultades y limitaciones. *Revista de Educación*, 1, 121-138. https://dialnet.unirioja.es/servlet/articulo?codigo=1332462

Perfetti, C. A. (1985). *Reading ability*. Oxford University Press.

Perfetti, C. A. & Adlof, S. M (2012). Reading comprehension: A conceptual framework from word meaning to text meaning. In J. P. Sabatini, E. Albro, & T. O'Reilly (Eds.), *Measuring up: Advances in how we assess reading ability* (pp. 3-20). Rowman & Littlefield Education.

Puente, A. (1991). *Comprension de la lectura y acción docente*. Editorial Pirámide.

Puente, A., Mendoza-Lira, M., Calderón, J. F., & Zúñiga, C. (2019). Estrategias metacognitivas lectoras para construir el significado y la representación de los textos escritos. *Ocnos*, *18*(1), 21-30. https://doi.org/10.18239/ocnos_2019.18.1.1781

Rasch, G. (1960). *Probabilistic models for asome intelligence and attainment tests* Nielsen & Lydiche, 184.

Rojas, L., Calderon, J. F., Ramírez, A., & Puente, A. (2024). Prueba de comprensión lectora para universitarios de nuevo ingreso: Análisis de las estrategias cognitivas y metacognitivas y su utilidad como variables predictoras del rendimiento académico. *Revista Iberoamericana de Diagnóstico y Evaluación Psicológica, 3*(73), 53-68. https://doi.org/10.21865/RIDEP73.3.04

Rojas, L., Puente, A., Gutiérrez-de-Blume, A., & Calderon, J. (2025a). Indicaciones sobre Prueba ELCU [Data set]. En *Ocnos*. Zenodo. https://doi.org/10.5281/zenodo.15537394

Rojas, L., Puente, A. Gutiérrez-de-Brume, A. & Calderon, J. F. (2025b). Procedimiento de uso plataforma SIGECOL para rendición prueba ELCU. [Data set]. En *Ocnos*. Zenodo. https://doi.org/10.5281/zenodo.15537422

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. https://doi.org/10.1191/0265532206lt337oa

Sabatini, J. P. (2002). Efficiency in word reading of adults: Ability group comparisons. *Scientific Studies of Reading*, 6, 267-298. https://doi.org/10.1207/S1532799XSSR0603_4

Sabatini, J. P. (2003). Word reading processes in adult learners. In E. M. H. Assink, & D. Sandra (Eds.), *Reading complex words: Cross-language studies* (pp. 265–94). London: Kluwer Academic. https://doi.org/10.1007/978-1-4757-3720-2_12

Schraw, G. (2009). A conceptual analysis of five measures of metacognitive monitoring. *Metacognition and Learning*, *4*(1), 33-45. https://doi.org/10.1007/s11409-008-9031-3

Schraw, G., & Moshman, D. (1995). Metacognitive theories. *Educational Psychology Review*, *7*(4), 351-371. https://doi.org/10.1007/BF02212307

Soto, C., Gutiérrez-de-Blume, A. P., Asun, R., Jacovina, M., & Vasquez, C. (2018). A deeper understanding of metacomprehension: Development of a new multidimensional tool. *Frontline Learning Research*, *6*(1), 31-52. https://doi.org/10.14786/flr.v6i1.328

Van-Dijk, T. A. (2006). Discourse, context, and cognition. *Discourse Studies, 8*(1), 159-177. https://doi.org/10.1177/1461445606059565

Van-Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension.* Academic Press.

Vega, N. Bañales, G., & Correa, S. (2012).¿Cómo los estudiantes universitarios autorregulan su comprensión cuando leen múltiples textos científicos? En *XI Congreso Nacional de Investigación Educativa*, México, D. F. (en línea). http://www.comie.org.mx/congreso/memoriaelectronica/v11/docs/area_01/2209.pdf

Yuill, N., & Oakhill, J. (1991). *Children's problems in text comprehension: An experimental investigation.* Cambridge University Press.