



---

## Evidence of content validity and internal consistency for a reading comprehension questionnaire administered to 10th-grade students

Evidencias de validez de contenido y consistencia interna de un cuestionario de comprensión lectora en estudiantes de décimo año EGB

Genny Maritza Reyes Pihuave\*  
José Luis Rodríguez Flores\*  
Carlos Fernando Ayala Quinto\*  
Luis Arturo Escobar Moreno\*

Received: January 21, 2026

Approved: March 12, 2026

Reyes, G., Rodríguez, J., Ayala, C., Escobar, L. (2026) Evidence of content validity and internal consistency for a reading comprehension questionnaire administered to 10th-grade students. *Espirales Revista Multidisciplinaria de investigación científica*, 10 (57), 32-

### Abstract

Reading comprehension is a key competence for school learning and requires technically supported instruments for its assessment. The objective of the study was to determine the evidence of content validity and internal consistency of a reading comprehension questionnaire for students in the tenth year of higher basic education. A quantitative, instrumental and descriptive study was developed. The instrument consisted of 20 open questions organized into literal, inferential and critical levels, assessed through an analytical rubric. The validity of the content was established with five expert judges and the internal consistency was estimated in a pilot test with 30 students. In addition, the questionnaire was applied as a pre-test and post-test to 76 students. The results reported Aiken's  $V = 0.96$ , Cronbach's  $\alpha = 0.906$ , and McDonald's  $\omega = 0.911$ . It is concluded that the instrument presents favorable initial evidence to evaluate reading comprehension in the context studied.

**Keywords:** reading comprehension; content validity; internal consistency; evaluation instrument; Higher Basic Education

---

\* Ministerio de Educación de Ecuador  
Guayaquil, Ecuador  
genny.reyes@docentes.educacion.edu.ec  
<https://orcid.org/0009-0004-6425-1228>

\* Universidad del Pacífico Facultad de  
Negocios y Economía Guayaquil, Ecuador  
<https://orcid.org/0000-0002-5475-0953>  
jose.rodriguez@upacifico.edu.ec

\* Ministerio de Educación de Ecuador  
Guayaquil, Ecuador  
ayala@docentes.educacion.edu.ec  
<https://orcid.org/0009-0008-2678-337X>

\* Ministerio de Educación de Ecuador  
Guayaquil, Ecuador  
<https://orcid.org/0009-0004-2393-8568>  
arturo.escobar@docentes.educacion.edu.ec

## Resumen

La comprensión lectora es una competencia clave para el aprendizaje escolar y exige instrumentos técnicamente sustentados para su evaluación. El objetivo del estudio fue determinar las evidencias de validez de contenido y consistencia interna de un cuestionario de comprensión lectora para estudiantes de décimo año de educación básica superior. Se desarrolló un estudio cuantitativo, instrumental y descriptivo. El instrumento estuvo conformado por 20 preguntas abiertas organizadas en los niveles literal, inferencial y crítico, valoradas mediante una rúbrica analítica. La validez de contenido se estableció con cinco jueces expertos y la consistencia interna se estimó en una prueba piloto con 30 estudiantes. Además, el cuestionario se aplicó como pretest y posttest a 76 estudiantes. Los resultados reportaron  $V$  de Aiken = 0,96, alfa de Cronbach = 0,906 y omega de McDonald = 0,911. Se concluye que el instrumento presenta evidencias iniciales favorables para evaluar comprensión lectora en el contexto estudiado.

**Palabras clave:** comprensión lectora; validez de contenido; consistencia interna; instrumento de evaluación; educación básica superior

## Introduction

Reading comprehension is a core skill for academic learning, as it goes beyond the decoding of written symbols to involve constructing meaning, identifying explicit information, making inferences, and forming judgments about texts. Its development has a direct impact on academic performance, participation in school life, and access to knowledge across various subject areas. International organizations have warned of the decline in reading skills following the pandemic and the need to strengthen evidence-based educational responses (UNESCO, 2021). Along the same lines, Duke et al. (2021) argue that reading comprehension should be understood as a multidimensional process that requires explicit instruction and rigorous assessment.

From a curricular perspective, reading comprehension is not a homogeneous skill but a progressive competency. In the Ecuadorian curriculum, reading performance is typically organized into literal, inferential, and critical levels, allowing for assessment ranging from the recognition of explicit information to the formulation of interpretations and reasoned judgments (Ministry of Education of Ecuador, 2016). Consequently, the quality of assessment depends not only on the pedagogical use of the instruments but also on the extent to which they are conceptually coherent and technically sound.

Recent research has emphasized that instruments designed to assess reading comprehension must provide explicit evidence of validity and reliability. Cervantes Buenfil and Canto Herrera (2025), in a systematic review of validated instruments for reading comprehension, found that psychometric support is a recurring requirement for validating the interpretation of scores and the educational decisions derived from them. Similarly, Trimiño-Pérez et al. (2024) note that many available reading tests have

weaknesses in their theoretical foundation or validation procedures, which limits their representational accuracy.

Another relevant aspect relates to the format of the items. Much of the recent literature continues to favor multiple-choice instruments, while open-ended formats remain less common, despite their potential to capture more complex comprehension processes. Cervantes Buenfil and Canto Herrera (2025) documented the predominance of closed-ended response measures in this field. However, evidence also shows that open-ended formats can achieve adequate technical quality when constructed based on a clear conceptual structure and accompanied by appropriate scoring procedures. Çelikgün et al. (2026), for example, reported favorable psychometric indicators for a reading comprehension and fluency assessment tool with open-ended questions. Furthermore, Menzala Peralta et al. (2024) argue that analytical rubrics improve the objectivity, feedback, and equity of assessment, which supports the use of open-ended items when seeking to examine complex reading performance.

In this context, the doctoral research from which this article derives designed a reading comprehension questionnaire with open-ended questions and an analytical rubric, organized around the literal, inferential, and critical levels. The instrument was aimed at tenth-grade students in upper secondary education and was administered in a real-world school setting. Although the original study examined the effect of an educational intervention, this manuscript shifts the focus to the assessment instrument itself. This decision is relevant because, as Sedlmayr and Weissenbacher (2025) note, the interpretation of reading scores depends not only on the construct but also on the conditions under which comprehension is measured.

Therefore, the purpose of this study was to determine the evidence of content validity and internal consistency of a reading comprehension questionnaire among tenth-grade students in upper secondary education. Consistent with this, the guiding question was as follows: What evidence of content validity and internal consistency does a questionnaire present for assessing reading comprehension at the literal, inferential, and critical levels among tenth-grade students in upper secondary education?

## Materials and methods

The study employed a quantitative approach and adopted an instrumental design with a descriptive scope, aimed at gathering evidence of content validity and internal consistency for a reading comprehension questionnaire. Although the instrument was originally used in a quasi-experimental study with pretest and posttest measurements, the focus of this article is not on the effectiveness of the educational intervention, but rather on the technical quality and empirical performance of the assessment instrument.

The research context was an educational institution in Guayaquil, Ecuador. In the original doctoral study, the population consisted of 152 students enrolled in the tenth grade of upper secondary education, distributed across four classes. The sample comprised 76 students selected through non-probabilistic purposive sampling and organized into two groups of 38 participants each. For the specific purposes of this

article, psychometric evidence for the instrument was obtained from three sources: the evaluation by five expert judges, a pilot test with 30 students, and the empirical administration of the questionnaire to the total sample at two measurement points.

The instrument consisted of a reading comprehension questionnaire comprising 20 open-ended questions. Responses were scored using an analytical rubric. The questionnaire was structured into three dimensions. The literal dimension included indicators related to text comprehension, the identification of explicit information, and the communication of information. The inferential dimension addressed inferential comprehension, logical deduction, and the interpretation of the meaning of words and phrases. The critical dimension included the evaluation of information, the analysis of arguments and opinions, and the formulation of judgments. The scoring system was organized with interpretation ranges both overall and by dimension. At the general level, scores between 20 and 40 indicated low performance, between 41 and 60 indicated average performance, and between 61 and 80 indicated high performance.

Content validity was established through expert judgment. Five specialists with doctoral degrees in education evaluated the questionnaire and rubric based on four criteria: adequacy, clarity, coherence, and relevance. The values obtained ranged from 0.92 to 1.00, and the overall coefficient estimated using Aiken's  $V$  was 0.96, indicating high agreement regarding the appropriateness of the content. However, some validation sheets included specific comments on the congruence of certain items with the corresponding indicators or dimensions, an aspect that should be understood as part of the normal technical review process for an educational instrument.

Internal consistency was estimated using a pilot test administered to 30 students. Reliability was calculated using Cronbach's  $\alpha$  and McDonald's  $\omega$ . The coefficients obtained were 0.906 and 0.911, respectively, which supports an adequate degree of homogeneity among the questionnaire items.

The procedure was carried out in four stages. In the first stage, the questionnaire and the analytical rubric were designed based on the theoretical and curricular framework of reading comprehension at the literal, inferential, and critical levels. In the second stage, the instrument was submitted to expert review for content validation. In the third stage, a pilot test was conducted to estimate internal consistency. In the fourth stage, the questionnaire was administered as a pretest and posttest to the sample of 76 students from the original study, allowing for the observation of its empirical performance under real school conditions.

Data analysis included Aiken's  $V$  for content validity and Cronbach's  $\alpha$  and McDonald's  $\omega$  for internal consistency. Statistical processing was performed using SPSS v25 and Jamovi. Additionally, descriptive statistics were used to organize the questionnaire results from the pretest and posttest applications, both globally and by dimension. In this article, these descriptive and inferential data are interpreted as complementary evidence of empirical functioning, rather than as the central purpose of the study.

Throughout the research process, the ethical principles of justice, autonomy, beneficence, and non-maleficence were upheld. Because the participants were minors, institutional authorization and informed consent from legal guardians were obtained prior to the pilot test and the school-based administrations of the instrument.

## Results

The first significant finding concerns the content validity of the questionnaire. The five expert raters evaluated the instrument in terms of adequacy, clarity, coherence, and relevance. Individual scores ranged from 0.92 to 1.00, and the overall Aiken's  $V$  reached 0.96. This result reflects a high level of agreement regarding the questionnaire's suitability as a measure of reading comprehension in the target population. Furthermore, the validation process showed that the instrument was, in general terms, well aligned with the dimensions it was intended to assess, although some specific items received comments aimed at semantic refinement or greater congruence with the corresponding indicator.

The second relevant result pertains to internal consistency. The pilot test with 30 students yielded a Cronbach's alpha of 0.906 and a McDonald's omega of 0.911. These coefficients indicate adequate internal consistency and suggest that the set of items functions cohesively with respect to the general construct of reading comprehension. The magnitude of both coefficients supports the preliminary reliability of the instrument for educational assessment purposes.

The administration of the pretest also provided evidence regarding the empirical performance of the questionnaire. At the overall reading comprehension level, the control group was distributed as follows: 31.6% at the low level and 68.4% at the medium level, with no students at the high level. The experimental group showed 57.9% at the low level and 42.1% at the medium level, also with no cases at the high level. This distribution indicates that the questionnaire was able to classify performance levels from the initial measurement.

By dimension, the questionnaire also captured distinct patterns. In the literal dimension, the control group had 26.3% at the low level, 63.2% at the medium level, and 10.5% at the high level, while the experimental group had 42.1% at the low level, 52.6% at the medium level, and 5.3% at the high level. In the inferential dimension, the control group recorded 39.5% at the low level, 55.3% at the medium level, and 5.3% at the high level, while the experimental group showed 63.2% at the low level and 36.8% at the medium level, with no cases at the high level. In the critical dimension, the control group recorded 47.4% at the low level and 52.6% at the medium level, while the experimental group had 76.3% of students at the low level, 21.1% at the medium level, and 2.6% at the high level. These results suggest that the instrument was sensitive to differentiated profiles across dimensions and groups.

In the post-test, the questionnaire again clearly classified performance levels. Overall, the experimental group had 65.79% of scores in the "meets learning objectives" category and 23.68% in "exceeds learning objectives," while the control group

remained primarily in “close to achieving learning” at 94.74%, and only 5.26% reached the “achieves learning” category. This pattern suggests that the instrument was capable of recording substantial changes in reading performance between the two measurement points.

At the dimensional level, the literal dimension showed 71.05% of the experimental group in “achieves learning” and 18.42% in “masters learning,” while the control group remained primarily in “close to achieving learning” at 89.47%. In the inferential dimension, the experimental group recorded 50.00% in “achieves learning” and 34.21% in “masters learning,” while the control group again concentrated 89.47% in “close to achieving learning.” In the critical dimension, the experimental group scored 55.26% in “achieves learning” and 26.32% in “masters learning,” while the control group remained mostly in “close to achieving learning” at 86.84%. These distributions provide empirical evidence that the questionnaire discriminated reading performance both overall and by dimension.

As complementary inferential evidence, the original study reported that there were no significant differences between the groups on the pretest, confirming baseline homogeneity. In contrast, the posttest showed statistically significant differences in the overall score, with  $t = -11.3$ ,  $p = 0.01$ , and an effect size of  $d = -2.59$  in favor of the experimental group. At the dimensional level, the posttest yielded  $U = 107$ ,  $Z = -6.43$ ,  $p = 0.000$ , and  $r = 0.74$  for the literal dimension;  $U = 199.5$ ,  $Z = -5.53$ ,  $p = 0.001$ , and  $r = 0.63$  for the inferential dimension; and a statistically significant advantage for the experimental group in the critical dimension, with a reported effect size of  $r = 0.67$ . Although these inferential data are not the focus of this article, they provide additional evidence that the questionnaire was able to capture distinct performance patterns in real educational settings.

The results of this study indicate that the reading comprehension questionnaire provides favorable evidence of content validity and internal consistency for tenth-grade students in upper secondary education. The overall Aiken’s  $V$  of 0.96 reflects a high degree of consensus among expert raters regarding the appropriateness of the instrument’s content, while the reliability coefficients ( $\alpha = 0.906$ ;  $\omega = 0.911$ ) support the internal consistency of the item set. In this regard, the study provides initial technical support for the use of the questionnaire in the educational context for which it was designed.

These findings align with research highlighting the importance of expert judgment and internal consistency in the development of educational instruments. Galicia Alarcón et al. (2017) argue that content validation through expert review is an essential procedure for determining whether an instrument adequately represents the variable it aims to measure. Similarly, Trimiño-Pérez et al. (2024) and Çelikgün et al. (2026) show that technically sound reading assessment instruments require conceptual clarity, systematic validation, and explicit evidence of reliability.

The results also align with recent studies on the validation of reading assessment tools. Trimiño-Pérez et al. (2024) reported favorable psychometric evidence for the Primary

Reading Proficiency Test, including inter-rater agreement and adequate internal consistency. Similarly, Çelikgün et al. (2026) found satisfactory alpha and omega coefficients in a reading comprehension and fluency assessment tool with open-ended questions. Although these studies were conducted with different populations and at different educational levels, they align with the present research in demonstrating that technically sound reading assessment instruments require conceptual clarity, systematic validation, and explicit evidence of reliability.

An additional strength of the present study lies in the use of open-ended questions accompanied by an analytical rubric. This feature is relevant because it allows for a more nuanced assessment of comprehension, including the retrieval of explicit information, the generation of inferences, and the formulation of judgments. Prah and Schuele (2022) found favorable evidence regarding construct validity for reading and listening comprehension methods that incorporated open-ended tasks at the passage level, suggesting that this type of format can capture complex comprehension processes. Likewise, Menzala Peralta et al. (2024) emphasize that rubrics improve grading objectivity, facilitate feedback, and promote equity in assessment. Consequently, the structure of the instrument used in this study is consistent with current arguments supporting more authentic and analytically rich forms of assessment.

However, the validation process also requires critical scrutiny. Although the overall content validity index was high, the expert review forms included observations regarding the wording and dimensional congruence of some items. This aspect is methodologically relevant because it shows that validation was not a mechanical confirmation of the questionnaire, but rather a technical review that identified aspects susceptible to refinement. This observation is consistent with Yeatman et al. (2024), who emphasize that the quality of reading assessment items depends on semantic precision, low ambiguity, and minimal contamination by demands for peripheral knowledge. Therefore, the results of the present study should be interpreted as evidence of strong overall adequacy, but not as proof of perfect performance of all items at the individual level.

A significant divergence emerges when comparing this study with the more psychometrically developed literature. While this research provides clear evidence of content validity and internal consistency, studies such as those by Trimiño-Pérez et al. (2024) and Çelikgün et al. (2026) extend the validation through exploratory or confirmatory factor analyses to examine the internal structure. Furthermore, the systematic review by Cervantes Buenfil and Canto Herrera (2025) shows that validated reading comprehension instruments typically report multiple forms of evidence, including construct, convergent, and predictive validity. In contrast, the present questionnaire still lacks these additional analyses. This does not invalidate the instrument, but it does indicate that its validation process remains incomplete and can be strengthened in future research.

Another point requiring caution relates to the interpretation of pretest and posttest results. In this article, the empirical performance of the questionnaire at two administration points should not be interpreted as evidence of test-retest stability, since

an educational intervention took place between the two measurements. Rather, these results indicate that the instrument was capable of classifying performance levels and recording changes under real-world school conditions. Sedlmayr and Weissenbacher (2025) note that reading assessment results are sensitive to administration conditions and the way the task is operationalized. Therefore, the inferences drawn from this study must remain tied to the specific context in which the questionnaire was administered.

Overall, the findings suggest that the main contribution of this study is not to establish comprehensive psychometric validation, but to offer initial and functionally useful technical evidence of a reading comprehension questionnaire designed for a specific school context. Its value lies in combining a conceptual structure aligned with the curriculum, open-ended questions, rubric-based scoring, expert judgment, and satisfactory internal consistency. Future studies should strengthen the questionnaire through internal structure analysis, convergent validity procedures, and applications in larger and more diverse samples.

.

## Conclusions

The reading comprehension questionnaire analyzed in this study provides strong evidence of content validity and internal consistency for use with tenth-grade students in upper secondary education. The expert review process revealed a high level of agreement regarding the adequacy, clarity, coherence, and relevance of the items, and the reliability coefficients supported the instrument's suitability for assessing reading comprehension in the study population.

The organization of the questionnaire into literal, inferential, and critical levels constitutes a methodological strength because it allows for the assessment of reading comprehension from a progressive and differentiated perspective. This structure makes it possible to distinguish between the retrieval of explicit information, the drawing of inferences, and the formulation of judgments.

The empirical application of the instrument in a school setting also showed that the questionnaire is effective for classifying performance levels and recording variations in reading comprehension both overall and by dimension. In this sense, the instrument not only has initial technical support but also empirical evidence of use in real-world assessment conditions.

However, these conclusions should be interpreted with caution. Although the results support content validity and internal consistency, additional studies are still needed to expand the psychometric evidence for the questionnaire through internal structure analysis, comparisons with other measures of reading comprehension, and applications in broader educational contexts. Therefore, the instrument can be considered a relevant tool for assessing reading comprehension in the studied population, but its future consolidation will require further validation processes.

## References

- Cervantes Buenfil, A. A., & Canto Herrera, P. J. (2025). *Análisis de instrumentos validados para evaluar la comprensión lectora en educación primaria: Una revisión sistemática (2015 a 2025)*. *RIDE Revista Iberoamericana para la Investigación y el Desarrollo Educativo*, 16(31). <https://doi.org/10.23913/ride.v16i31.2633>
- Çelikgün, B., Akdaş, F., Eser, B. N., Minga, R., Ölçek, G., Gücüyener, C., & Edman, S. (2026). *Development and validation of a reading comprehension and fluency screening assessment tool for children aged 7–10: Implications for audiological rehabilitation*. *Frontiers in Psychology*, 17, 1759333. <https://doi.org/10.3389/fpsyg.2026.1759333>
- Duke, N. K., Ward, A. E., & Pearson, P. D. (2021). The science of reading comprehension instruction. *The Reading Teacher*, 74(6), 663–672. <https://doi.org/10.1002/trtr.1993>
- Galicia Alarcón, L. A., Balderrama Trápaga, J. A., & Edel Navarro, R. (2017). Validez de contenido por juicio de expertos: Propuesta de una herramienta virtual. *Apertura*, 9(2), 42–53. <https://doi.org/10.32870/ap.v9n2.993>
- Menzala Peralta, R. M., Ortega Menzala, E., & Zanabria Vargas, E. (2024). Uso de la rúbrica en la educación: Una revisión sistemática. *Horizontes. Revista de*

*Investigación en Ciencias de la Educación*, 8(34), 1727–1743.

Ministerio de Educación del Ecuador. (2016). *Currículo de los niveles de educación obligatoria*. Ministerio de Educación.

Prahl, A., & Schuele, C. M. (2022). A pilot study assessing listening comprehension and reading comprehension in children with Down syndrome: Construct validity from a multi-method perspective. *Frontiers in Psychology*, 13, 905273.

<https://doi.org/10.3389/fpsyg.2022.905273>

Sedlmayr, P., & Weissenbacher, B. (2025). Reading comprehension assessment for student selection: Advantages of text availability in terms of validity. *Frontiers in Education*, 10, 1524561. <https://doi.org/10.3389/feduc.2025.1524561>

Trimiño-Pérez, L., Hurtado-Reina, J., Velasco, E., & Martínez, A. (2024). Design and validation of a Primary Reading Proficiency Test (PCL-P). *Frontiers in Language Sciences*, 3, 1471040.

<https://doi.org/10.3389/flang.2024.1471040>

UNESCO. (2021, 26 de marzo). *Cien millones más de niños sin las competencias mínimas de lectura debido a la COVID-19: La UNESCO reúne a los ministros de educación*. UNESCO.

Yeatman, J. D., Tran, J. E., Burkhardt, A. K., Ma, W. A., Mitchell, J. L., Yablonski, M., Gijbels, L., Townley-Flores, C., & Richie-Halford, A. (2024). Development and validation of a rapid and precise online sentence reading efficiency

assessment. *Frontiers in Education*,  
9, 1494431.  
<https://doi.org/10.3389/feduc.2024.1494431>