

<https://doi.org/10.23913/ride.v16i31.2633>

*Artículos científicos*

## **Análisis de instrumentos validados para evaluar la comprensión lectora en educación primaria: Una revisión sistemática (2015 a 2025)**

*Analysis of validated instruments for assessing reading comprehension in primary education: A systematic review (2015 to 2025)*

*Análise de instrumentos validados para avaliação da compreensão de leitura no ensino fundamental: uma revisão sistemática (2015 a 2025)*

**Adela Aurora Cervantes Buenfil**

Universidad Autónoma de Yucatán, México

[adeceba28@gmail.com](mailto:adeceba28@gmail.com)

<https://orcid.org/0009-0003-2472-7204>

**Pedro José Canto Herrera**

Universidad Autónoma de Yucatán, México

[pcanto@correo.uady.mx](mailto:pcanto@correo.uady.mx)

<https://orcid.org/0000-0001-5428-8343>

### **Resumen**

La comprensión lectora representa una habilidad cognitiva que repercute en el rendimiento académico de los estudiantes y que requiere ser evaluada con instrumentos de alta confiabilidad. Por tales motivos, se llevó a cabo una revisión sistemática con el objetivo de identificar instrumentos validados para evaluar la comprensión lectora en educación primaria y analizar sus características de diseño y propiedades psicométricas. La metodología implementada fue mediante la búsqueda de artículos empíricos en las bases de datos EBSCO, Scopus y Web of Science, con un período comprendido de 2015 a 2025. Finalmente, se obtuvo un corpus de 28 artículos que cumplieron con los criterios de inclusión y exclusión, ya que son instrumentos con propiedades psicométricas. El análisis de estos reveló que predominan pruebas con ítems de opción múltiple que se basan en la lectura de textos



narrativos y expositivos, con niveles adecuados y altos de confiabilidad y con diversas formas de validez: contenido, constructo, convergente y predictivo. Así como un conjunto de pruebas de ejecución máxima que se concentran en examinar determinados procesos cognitivos y metacognitivos. No obstante, aún no se dispone de instrumentos que evalúen varios procesos cognitivos y lingüísticos de manera integrada y se sugiere ampliar los estudios sobre validez predictiva y sensibilidad al cambio.

**Palabras clave:** comprensión lectora, evaluación, lectura, primaria.

### **Abstract**

Reading comprehension represents a cognitive skill that affects students' academic performance and must be assessed using highly reliable instruments. For this reason, a systematic review was carried out to identify validated tools for evaluating reading comprehension in primary education and to analyze their design features and psychometric properties. The methodology involved searching for empirical studies in the EBSCO, Scopus, and Web of Science databases over a ten-year period (2015–2025). A corpus of 28 articles that met the inclusion and exclusion criteria was obtained, each reporting an instrument for measuring reading comprehension with documented psychometric properties. Analysis revealed a predominance of multiple-choice tests based on narrative and expository texts, exhibiting adequate to high reliability and various forms of validity (content, construct, convergent, and predictive). In addition, a set of performance-based tasks was identified that focused on examining specific cognitive and metacognitive processes. However, there remains a lack of instruments that assess multiple cognitive processes in an integrated fashion, and future research should expand on predictive validity and sensitivity to change studies.

**Keywords:** reading comprehension, assessment, reading, primary education.

## Resumo

A compreensão leitora representa uma habilidade cognitiva que impacta o desempenho acadêmico dos alunos e requer avaliação com instrumentos altamente confiáveis. Por essas razões, uma revisão sistemática foi conduzida para identificar instrumentos validados para avaliar a compreensão leitora no ensino fundamental e analisar suas características de delineamento e propriedades psicométricas. A metodologia implementada envolveu a busca de artigos empíricos nas bases de dados EBSCO, Scopus e Web of Science, abrangendo o período de 2015 a 2025. Ao final, foi obtido um corpus de 28 artigos que atenderam aos critérios de inclusão e exclusão, por serem instrumentos com propriedades psicométricas. A análise desses artigos revelou uma predominância de testes com itens de múltipla escolha baseados na leitura de textos narrativos e expositivos, com níveis adequados e altos de confiabilidade e diversas formas de validade: de conteúdo, de construto, convergente e preditiva. Isso também é acompanhado por um conjunto de testes de desempenho máximo que se concentram em examinar processos cognitivos e metacognitivos específicos. No entanto, ainda não há instrumentos disponíveis que avaliem vários processos cognitivos e linguísticos de forma integrada, e sugere-se que mais estudos sobre validade preditiva e sensibilidade à mudança sejam realizados.

**Palavras-chave:** compreensão leitora, avaliação, leitura, ensino fundamental.

**Fecha Recepción:** Mayo 2025

**Fecha Aceptación:** Septiembre 2025

---

## Introducción

La comprensión lectora (CL) es un proceso clave para el desarrollo integral de las niñas y niños, ya que favorece la consolidación de conocimientos, la comunicación interpersonal y la participación activa en la sociedad (Avendaño-Guevara y Balderas-Gutiérrez, 2023). Cabe señalar que, durante la educación primaria, esta habilidad es la base para tener un desempeño académico exitoso y, en etapas posteriores, incide en las oportunidades laborales e inclusión social (Esmer y Güneş, 2019; Tamay-Loaiza, 2024; Thomas *et al.*, 2024). Sin embargo, el bajo desempeño en CL es un fenómeno que persiste y que se ha perpetuado, que lejos de aminorarse, tiende a agudizarse (Gallegos, 2024).

Esta problemática presenta impacto a nivel mundial, ya que se estima que el 70 % de las niñas y niños de 10 años de edad en países de medios y bajos ingresos no logran comprender textos de baja dificultad (UNICEF, 2022). Específicamente, en Latinoamérica y el Caribe, cuatro de cada cinco menores de edad no pueden leer y comprender textos



simples, lo cual implica que, al término de la educación primaria, estos estudiantes carecerán de las competencias mínimas para la CL (UNICEF, 2022).

Cabe señalar que tras la pandemia COVID-19, estos resultados se intensificaron, es decir, la falta de CL se agudizó (Fuchs *et al.*, 2023). En consecuencia, existe la necesidad urgente de poner en marcha pruebas para medirla objetivamente, que coadyuven a la planificación de cursos de nivelación que aborden las dificultades de aprendizaje en conocimientos básicos de educación primaria (Hevia *et al.*, 2022).

Ante estas circunstancias, uno de los primeros pasos para atender esta problemática es aplicar instrumentos de evaluación para determinar el nivel de desempeño de los estudiantes. Estos resultados deben ser el punto de partida y no invisibilizar las necesidades educativas. Tal como señala Lozada *et al.* (2016), la evaluación es un pilar para mejorar la calidad de los aprendizajes, en especial en áreas como la CL, por lo que se requiere seguir generando formas de evaluación que proporcionen información del rendimiento académico.

No obstante, queda claro que medir la CL no resulta tarea fácil, ya que no es un proceso unitario, es decir, puede descomponerse en varios procesos, operacionalizarse y medirse centrándose en alguno de estos (Moncada *et al.*, 2024). En consecuencia, es especialmente importante que los educadores, investigadores e incluso creadores de materiales didácticos implementen instrumentos precisos y fiables para medir la CL. Los resultados no solamente son indicadores de esta habilidad, sino que son herramientas esenciales para determinar el rendimiento de los estudiantes (Esmer y Güneş, 2019).

Por lo tanto, la utilidad de contar con instrumentos de evaluación bien diseñados, confiables y validados, radica en la posibilidad de ofrecer datos verificados que orienten políticas educativas y prácticas pedagógicas en el aula, principalmente en los primeros grados escolares, debido a que es la etapa decisiva que sienta las bases para las habilidades lectoras (Lozada *et al.*, 2016 y Míguez-Álvarez *et al.*, 2022).

En el ámbito de la investigación educativa, en el género de las revisiones sistemáticas, se han encontrado artículos que abordan diversas problemáticas relacionadas con la CL y diferentes tipos de intervenciones pedagógicas. Por ejemplo, Díaz Calle *et al.* (2024) se orientan al análisis de la CL en la era digital, enfocándose en realizar una revisión de estrategias innovadoras y que emplean tecnologías educativas que fomentan esta habilidad.

Álvarez Porroa *et al.* (2022) desarrollaron una revisión sistemática acerca de la CL en la etapa escolar, en la cual analizan los conceptos inherentes a esta habilidad. En el caso de Chávez-Delgado *et al.* (2022) presentan una revisión sistemática centrada en programas de

intervención en lectura inicial, por lo que evidencia cómo las intervenciones sistemáticas y estructuradas benefician el aprendizaje lector de primero de primaria. Asimismo, Austin *et al.* (2019) muestran la relación entre la calidad metodológica de los estudios y el efecto de las intervenciones suplementarias en lectura en estudiantes con dificultades o en riesgo. Por último, el trabajo de Ramírez Cubas *et al.* (2025) que abarca una revisión sistemática de CL en estudiantes de educación básica, por lo que se centran en analizar investigaciones actuales en Perú particularmente.

En respuesta a esta problemática, la presente revisión sistemática (2015-2025) se propone como objetivo: identificar instrumentos validados para evaluar la CL en educación primaria y analizar sus características de diseño, es decir, formato de ítems, modalidad de aplicación y niveles de comprensión que miden.

## Método

El presente estudio corresponde a una revisión sistemática de la literatura, la cual comprende un resumen claro y estructurado de la información disponible, que se encuentra orientada a cubrir el objetivo previamente planteado. Cabe señalar que se constituye por múltiples artículos empíricos, que conforman un alto nivel de evidencia (Moreno *et al.*, 2019). Es importante mencionar que está basada en la lista de criterios de la declaración PRISMA (*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*) (Page *et al.*, 2021).

## Estrategias de búsqueda

Los artículos empíricos analizados para la presente revisión sistemática fueron seleccionados en revistas científicas indexadas a bases de datos de alto impacto, específicamente Web of Science, Scopus y EBSCO. Los años de las publicaciones abarcan desde 2015 hasta 2025. Asimismo, para los fines de la revisión sistemática, fue necesario plantear la pregunta de investigación: ¿Qué instrumentos validados miden la comprensión lectora en educación primaria y cuáles son sus características de diseño?

Por su parte, las palabras clave de la investigación y las cadenas de búsqueda se diseñaron con la estrategia PEO (Bettany-Saltikov, 2012), por lo que se consideraron los tres elementos principales: Población, Exposición y Resultados (véase tabla 1).

**Tabla 1.** Pregunta de investigación y palabras clave diseñadas con PEO.

Pregunta de investigación ¿Qué instrumentos validados miden la comprensión lectora en estudiantes de primaria y cuáles son sus características de diseño?		
Población (1)	Exposición (2)	Resultados (3)
Estudiantes de primaria Estudiantes Educación primaria	Instrumentos validados para medir la comprensión lectora. Prueba de comprensión lectora	Diseño de instrumento Formato de ítems Modalidad de aplicación Niveles evaluados
Primary school student Pupils Primary education	Validated instruments Reading comprehension test Instrument / assessment	Instrument design Item format Administration mode Assessed comprehension levels

Fuente: Elaboración propia

### Cadenas de búsqueda

Para llevar a cabo la consulta en las diferentes bases de datos, se emplearon cadenas de búsqueda conformadas por las palabras clave y marcadores booleanos AND y OR. La ecuación principal se representa en la siguiente tabla (véase tabla 2).

**Tabla 2.** Cadena de búsqueda para las bases de datos

Cadena de búsqueda
<i>reading comprehension AND (primary OR "primary education") AND (student OR pupil) AND (instrument* OR test* OR scale* OR questionnaire*) AND (valid* OR reliab* OR psychometric*)</i>

Fuente: Elaboración propia.

### Criterios de inclusión y exclusión.

Se analizaron los criterios de inclusión y exclusión, con el objetivo de identificar artículos empíricos que cumplieran con el propósito de la revisión sistemática y aquellos que presentan características que no contribuyen al estudio. En consecuencia, se empleó la tabla PEO como guía (Bettany-Saltikov, 2012) (véase tabla 3).

**Tabla 3.** Criterios de inclusión y criterios de exclusión

	Criterios de inclusión	Criterios de exclusión
Población	Estudiantes de primaria	Niños de preescolar, secundaria, educación media superior o superior Docentes, padres de familia.
Exposición	Instrumentos de evaluación validados que midan la CL.	Instrumentos que no hayan sido validados o no sean para CL.
Resultado	Investigaciones que describan características del diseño del instrumento y su aplicación.	Publicaciones que sólo mencionen el uso de pruebas sin incluir ningún indicador de calidad o sin resultados claros sobre su impacto en la evaluación/mejora de la CL.
Tipo de estudio	Artículos empíricos	Revisión de literatura, capítulos de libro, conferencias, ensayos, tesis, ponencias
<b>Filtros</b>		
Idiomas	Español Inglés	Otros
Años	2015-2025	

Fuente: Elaboración propia.

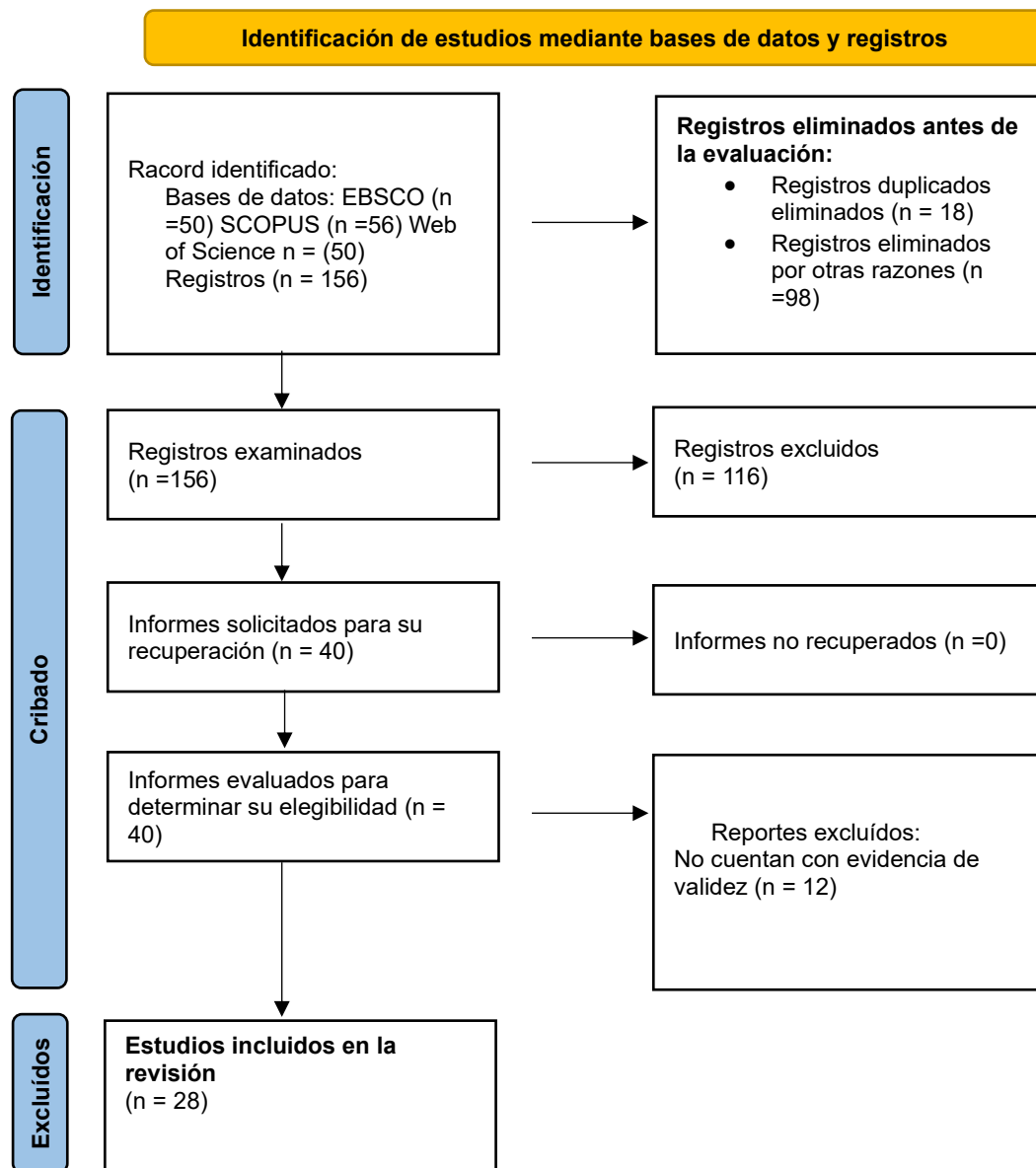
### Selección de estudios

La selección de los estudios que contribuyen a resolver la pregunta de la revisión sistemática, fue llevada a cabo por la implementación de diferentes fases de búsqueda. La primera fase consistió en leer los títulos, palabras clave y abstract en las diferentes bases de datos y por medio de los criterios de inclusión y exclusión. En consecuencia, se tuvieron los siguientes resultados: EBSCO 50 artículos, Scopus 56 y Web of Science 50 estudios empíricos; lo que da un total de 156 artículos científicos.

La segunda fase consistió en la lectura completa de las investigaciones seleccionadas, de modo que se leyeron con detenimiento las características de los instrumentos utilizados en cada investigación. Además, la selección requirió de tener en cuenta tres aspectos: a) Usar pruebas de ejecución máxima, b) Contar con validez de contenido, criterio o constructo

y c) Describir el instrumento empleado para la medición. Finalmente, se obtuvo un corpus consistente con los CI-CE: EBSCO 12, Scopus ocho artículos científicos y Web of Science ocho; con un total de 28 artículos para desarrollar la lectura completa y a detalle (fig. 1).

**Figura 1.** Diagrama de flujo basado en PRISMA



Fuente: Diagrama de flujo tomado del protocolo de la declaración PRISMA 2020 (Page *et al.*, 2021).

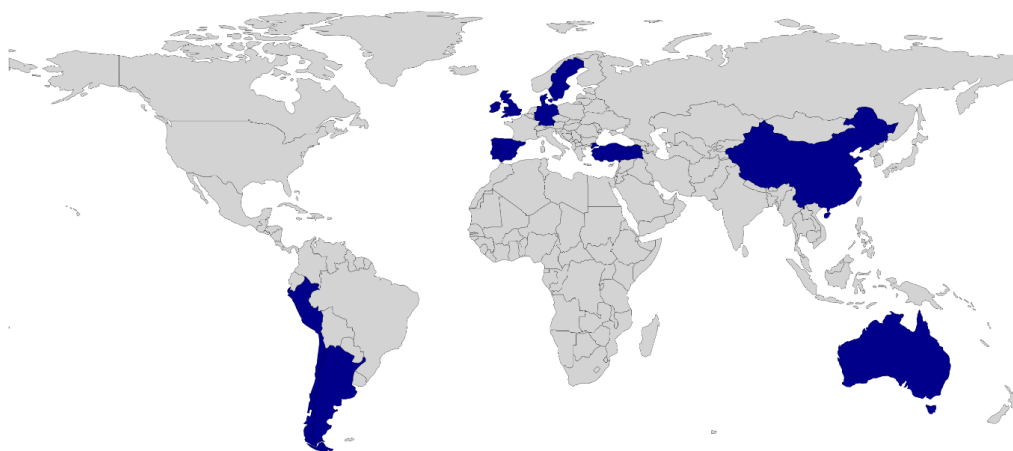
## Análisis de datos

La sistematización de la información más importante de los artículos empíricos, requirió la elaboración de una matriz de Excel con datos de los estudios de interés. Entre los datos sistematizados se encuentran nombre del artículo, nombre de los autores, resumen, doi, base de datos de origen, año del estudio, país de procedencia, tipo de investigación, tamaño de la muestra, aportaciones y nivel educativo primaria.

## Resultados

Se obtuvo un total de 28 artículos recopilados y analizados de los cuales 25 % (n=7) pertenecen a España, 14 % (n=4) a Alemania, 7,1 % (n=2) a Portugal y 7,1 % (n=2) a Escocia. Asimismo, podemos encontrar varios países europeos con el 3,5 % (n=1) es decir un artículo cada país, entre estos Dinamarca, Irlanda y Suecia. En Australia y China solamente fue hallado un artículo. Por otro lado, en el país euroasiático, representado por Turquía se observa 7,1 % (n=2) de todo el corpus de artículos con instrumentos de CL validados. Finalmente, en los países latinoamericanos sobresale Chile 10,7 % (n=3) de los estudios y Argentina y Perú con un artículo, es decir, con 3,5 % (n=1) (fig. 2).

**Figura 2.** Mapa con la distribución geográfica de los artículos encontrados.



Fuente: Elaboración propia.

A continuación, se presentan las principales temáticas y resultados de la literatura revisada en tres apartados, los cuales son los siguientes: fundamentos teóricos del diseño de instrumentos, tipos y estructura de los instrumentos y propiedades psicométricas de los instrumentos de CL.

## Fundamentos teóricos del diseño de instrumentos.

Para llevar a cabo la aplicación de instrumentos de evaluación para la CL, los diseñadores de evaluaciones requieren conocer a profundidad la teoría que fundamenta el instrumento que implementan, para asegurar su fiabilidad (Schindler *et al.*, 2018; Tennent, 2020). De manera que en los siguientes párrafos se analizan las bases teóricas que sustentan los artículos empíricos del corpus presente.

De acuerdo con Barreyro *et al.* (2020), Marín-Marín *et al.* (2023) y Seifert y Paleczek, (2022) la CL se entiende como la habilidad de decodificar un texto y de construir su significado, a través de la interacción del lector con la información del texto y la integración con sus conocimientos y experiencias previas. Esta manera de conceptualizar la CL tiene raíces en el enfoque sociocultural, lo que lleva a darle un valor importante al contexto de los estudiantes, ya que esto indica que los lectores interactúan con la información por lo que su comprensión estará mediada por múltiples factores personales, los cuales son indispensables reconocer para llevar a cabo mediciones eficientes (Li, 2022; Tennent, 2020)

Por otro lado, autores como Barreyro *et al.* (2020), Figueroa-Sepúlveda y Gallego-Ortega (2018), Jungjohann *et al.* (2018), Marín-Marín *et al.* (2023), Míguez-Álvarez *et al.* (2022), Pascual *et al.* (2021) y Schindler *et al.* (2018), consideran que la CL es una habilidad de la que subyacen varios procesos cognitivos complejos como la denominación automática, la memoria de trabajo, las funciones ejecutivas, la sintaxis, fluidez lectora, la fonética, el léxico, la decodificación y los procesos metacognitivos, por lo que los autores evalúan estos procesos para identificar su relevancia y relación con la CL.

Con relación a los procesos metacognitivos estos son considerados esenciales al evaluar la CL. Tal como afirman Barreyro *et al.* (2020), Marín-Marín *et al.* (2023), Pascual *et al.* (2021), Quispe-Morales, (2022) y Ronqui *et al.* (2024), en las pruebas que implementan se examinan procesos como la autorregulación, el monitoreo de la CL e incluso habilidades para detectar errores como pseudopalabras, o el procedimiento tipo *Cloze* que le indica al estudiante llenar espacios vacíos en oraciones o textos con incongruencias semánticas.

Otro referente teórico es el modelo de construcción integración, debido a que este modelo es la representación mental más profunda que construye el lector al combinar la información explícita del texto con sus conocimientos y experiencias previas, Van Dijk y Kintsch (1983, citado en Barreyro *et al.*, 2020; Schindler *et al.*, 2018; Figueroa-Sepúlveda & Gallego-Ortega, 2018; Jungjohann *et al.*, 2018).

De acuerdo con este modelo la CL va en progreso gradualmente, por lo que parte de la decodificación de palabras, es decir, este es un nivel superficial, posteriormente se organiza en oraciones y finalmente se crean inferencias e imágenes mentales más complejas. Barreyro *et al.* (2020) aprovechan estos niveles insertando errores intencionales en diferentes unidades textuales para evaluar cómo los estudiantes detectan y corrigen incongruencias en cada fase. Schindler *et al.* (2018) diseñan tareas de juicio de gramaticalidad que validan la transición de procesos literales a inferenciales mediante modelos de respuesta al ítem, estructuran su material multimedia en actividades que van de preguntas literales con palabras, a estructurales con oraciones y finalmente inferenciales con textos, ya que miden las estrategias metacognitivas asociadas a cada nivel.

De modo similar, Figueroa-Sepúlveda y Gallego-Ortega (2018) implementan el CLP para diferenciar respuestas locales, globales e inferenciales en alumnos de primaria, mientras que Jungjohann *et al.* (2018) se centran explícitamente en la comprensión a nivel de oración, puesto que evalúan la capacidad de los lectores para procesar y juzgar la coherencia y corrección sintáctica en enunciados.

### **Tipos y estructura de los instrumentos.**

Los instrumentos validados para evaluar CL a nivel de educación primaria, comparten características inherentes a su diseño y aplicación, en específico 53,5 % (n= 15) de las investigaciones implementó evaluaciones con pasajes textuales narrativos y expositivos, que posteriormente se acompañan de preguntas de opción múltiple con las que se evalúan niveles de CL como el literal, inferencial y crítico.

En la categoría antes descrita, destacan instrumentos como la Prueba de Comprensión Lectora de Complejidad Progresiva (CLP) aplicada en los artículos de Barreyro *et al.* (2020) y Figueroa-Sepúlveda y Gallego-Ortega (2018); CLIP o Test de Comprensión Lectora para Intervención en Primaria implementado por Pascual *et al.* (2021); la Evaluación de la Comprensión Lectora para Alumnos de Primaria (ECOMPLETEC-Pri) citada en Míguez-Álvarez *et al.* (2022); el instrumento alemán *Grazer Leseverständnistest* (GraLeV) empleado por Seifert & Paleczek, (2022); ELFE II que es un test de CL igualmente de Alemania de acuerdo con Visser *et al.* (2023); la prueba STAR Reading citada en Topping (2018); ACL-3 instrumento seleccionado para el estudio de Martínez-Álvarez *et al.* (2020); LegiLexi (Nordström *et al.*, 2025); el Test Objetivo Castellano (Quispe-Morales, 2022); en China la evaluación DCRCA de Li *et al.* (2021); TCL Test de comprensión lectora empleado por

Cadime *et al.* (2015), el instrumento británico WIAT-IIUK-T RCS de McCartney *et al.* (2015) y Woodcock-Johnson Test de O'Rourke *et al.* (2016). Aunado a lo antes mencionado, dichos instrumentos son aplicados de manera individual o grupal, adaptándose tanto a formatos impresos como digitales.

Estos mismos instrumentos, presentan particularidades como es el caso de la prueba ACL-3 específicamente para tercer grado de primaria, que incluye siete textos con 25 ítems y evalúa la comprensión literal, reorganización, el nivel inferencial y crítico (Martínez-Álvarez *et al.*, 2020). De manera similar, Cadime *et al.* (2015), se orienta a la evaluación de estos mismos procesos de la CL, pero en segundo, tercero y cuarto de primaria y mediante textos originales que incluyen instructivos y poemas.

En el caso del test denominado CLIP este puede ser implementado de 3° hasta 6° grado de primaria, ya que cuenta con dos versiones la primera con textos cortos y la segunda con textos de mayor amplitud, en ambos niveles son de tipo narrativos y expositivos (Pascual *et al.*, 2021). De acuerdo con Barreyro *et al.* (2020) y Figueroa Sepúlveda *et al.* (2018), la prueba CLP se encuentra validada y estandarizada en el contexto chileno y es aplicada en alumnos de 3° y 4° grado de primaria, mediante la lectura de textos con ocho niveles de dificultad.

Se han analizado pruebas digitales como es el caso de Li *et al.* (2021) que al utilizar la evaluación DCRCA de China da a conocer que su diseño abarca de 2° a 6° de primaria cuenta con 48 ítems, divididos en tres cuadernillos, con 16 preguntas y evalúa por medio de una matriz que mide con precisión las habilidades que domina el estudiante y genera un reporte inmediato con las fortalezas y áreas de oportunidad. Asimismo, ECOMPLEC-Pri usado por Míguez-Álvarez *et al.* (2022) cuenta con 22 preguntas de opción múltiple después de la lectura de textos largos y sus respuestas son evaluadas a través de un software que genera puntuaciones globales e individuales. Por su parte, LegiLexi de Nordström *et al.* (2025), es un instrumento digitalizado sueco, que emplea textos breves apoyados con imágenes que seleccionan las niñas y niños para responder los ítems.

Sumado a lo anterior, se identifican las pruebas ELFE y GraLev con dos versiones, ya sea digital o en papel y que exploran la CL desde el análisis a nivel palabra, oración y texto (Seifert & Paleczek, 2022; Visser *et al.*, 2023). Míguez-Álvarez *et al.* (2022) y Ramos-Carrillo y Moreno-Pérez, (2023), coinciden al seleccionar PROLEC-R, instrumento con formatos mixtos que combina preguntas abiertas e ítems cerrados al aplicarse a los alumnos,

está conformado por una etapa individual y una grupal permiten captar evaluaciones más amplias y profundas, especialmente en comprensión inferencial y crítica (tabla 4).

**Tabla 4.** Lista de instrumentos con textos narrativos y expositivos.

Instrumento Cita del artículo	Descripción / Ítems	Aplicación/ Nivel de primaria.	Formato	Proceso evaluado
ACL-3 (Martínez-Álvarez <i>et al.</i> , 2020)	7 textos de 25 ítems	Grupal/3 <sup>o</sup>	Opción múltiple	Cuatro niveles de comprensión: literal, reorganización, inferencial y crítica.
CLIP (Test de Comprensión Lectora para Intervención en Primaria) (Pascual <i>et al.</i> , 2021)	66 ítems totales: textos cortos (36 ítems) y largos (30 ítems).	Grupal/3 <sup>o</sup> a 6 <sup>o</sup> de primaria	Opción múltiple	Comprensión local, global e inferencial
CLP (Complejidad Progresiva) (Barreyro <i>et al.</i> , 2020 ; Figueroa- Sepúlveda <i>et al.</i> , 2018).	8 niveles de dificultad, Se implementan textos narrativos y expositivos. Cada nivel con varios textos.	Grupal/3 <sup>o</sup> y 4 <sup>o</sup>	Opción múltiple	Comprensión global
DCRCA (Li <i>et al.</i> , 2021)	3 cuadernillos (16 ítems c/u).	Digital/2 <sup>o</sup> a 6 <sup>o</sup> grado	Opción múltiple	Diagnóstico de subhabilidades lectoras cognitivas: recuperación de información, inferencias, integración y síntesis, vocabulario, nivel crítico y tipo de textos.
ECOMPLEC-Pri Evaluación de la Comprensión Lectora para primaria (Míguez-Álvarez <i>et al.</i> , 2022)	21 ítems	Oral / digital en 4 <sup>o</sup> 5 <sup>o</sup> y 6 <sup>o</sup> grado.	Opción múltiple	Literal e inferencial

ELFE II Test de Comprensión Lectora 2ª edición. (Visser <i>et al.</i> , 2023)	Mide 3 niveles: palabra (75 ítems), oración (36 ítems), texto (26 ítems).	Digital / papel Duración de 20 a 30 minutos. 1º a 6º grado	Opción múltiple	Evalúa a nivel palabra, oraciones y textos.
GraLeV (Seifert & Paleczek 2022)	4 subpruebas: a nivel palabra evalúan 12 palabras Nivel oración (16 ítems con imagen), Texto I (8 textos breves, 2 preguntas), Texto II (2 textos, 15 preguntas).	Digital / impreso en 3º y 4º grado.	Opción múltiple	Evalúa a nivel palabra, oraciones y textos.
LegiLexi (Nordström <i>et al.</i> , 2025)	12 textos breves con ilustraciones para seleccionar acuerdo con el texto, 6 textos largos.	En Tablet para alumnos de 1º a 3º.	Opción múltiple	Literal, global e inferencial
PROLEC-R (Míguez-Álvarez <i>et al.</i> , 2022; Ramos-Carrillo & Moreno-Pérez, 2023)	4 pasajes (2 cortos, 2 largos), 16 preguntas abiertas inferenciales.	Individual en 4º 5º y 6º grado	Abiertas	Inferencial
Test Objetivo Castellano (Quispe-Morales, 2022)	30 ítems (3 bloques de 10 preguntas):	Grupal 6º grado	Opción única	Miden nivel literal, inferencial y crítico.
TCL Test de comprensión lectora (Cadime <i>et al.</i> , 2015)	30 ítems, se basa en la lectura de textos literarios.	Grupal 3º 4º y 5º grado.	Opción múltiple	Comprensión literal inferencial reorganización y crítica.
WIAT-IIUK-T RCS (McCartney <i>et al.</i> , 2015)	Lectura de pasajes textuales narrativos y expositivos.	Individual 3º y 4º	Preguntas Abiertas y opción múltiple.	Comprensión oral y escrita

Woodcock-Johnson Tests (O'Rourke <i>et al.</i> , 2016)	74 ítems	Individual Estudiantes de 8 años	Lectura de pasajes textuales.	Cuatro niveles de comprensión: literal, reorganización, inferencial y crítica.
---	----------	-------------------------------------	-------------------------------	--

Fuente: Elaboración propia.

Otro diseño instrumental recurrente implica técnicas tipo *Cloze*, en las cuales los estudiantes completan espacios en blanco dentro de un texto, ya que evalúan la comprensión local y global, es decir la comprensión textual y aquella que requiere del dominio de inferencias y predicciones. Entre estos, se halla la prueba TORCH citada por Hay & Stevenson, (2024); HyTeC-Cloze en Moncada *et al.* (2024) y CBM-Maze de oraciones, prueba alemana identificada en la investigación de Jungjohann *et al.* (2018) que ejemplifican claramente esta técnica caracterizada por la aplicación digital.

Cabe señalar que estas pruebas son aplicadas a partir de tercer grado de primaria, en particular, TORCH es de tipo grupal y se apoya de pasajes textuales, en este conjunto de pruebas es la única que se administra en papel. En contraste, CBM-Maze, que es en línea y que orienta su medición al análisis de oraciones verificando la sintaxis y semántica; lo que representa una opción más rápida y práctica si se cuenta con los recursos propicios. Con referencia a HyTeC-Cloze, también es una herramienta en línea, sin embargo, es una prueba de mayor extensión: consta de 18 textos de mediana extensión que abordan contenidos curriculares (tabla 5).

**Tabla 5.** Lista de instrumentos con espacios en blanco.

Instrumento Autor(es)	Descripción / Ítems	Aplicación/ Nivel de primaria.	Proceso evaluado
CBM-Maze de oraciones (Jungjohann <i>et al.</i> , 2018)	Prueba <i>Cloze</i> con distractores. Se aplica en 8 minutos en línea	Digital para 3°.	CL a nivel de oración.
HyTeC-Cloze en español. (Moncada <i>et al.</i> , 2024)	18 textos y 2 versiones; 20–30 espacios en blanco; sinónimos permitidos.	Digital para 6°.	Comprensión local y global
TORCH (Hay & Stevenson, 2024)	Pasajes sin límite de tiempo. Varias páginas con espacios a completar.	Grupal 3° a 6° grado	CL e inferencias.

Fuente: Elaboración propia.

Asimismo, existe un énfasis especial en instrumentos que evalúan procesos metacognitivos y de autorregulación durante la lectura y que igualmente evalúan la CL, entre estos sobresalen REMICOM citado en Barreyro *et al.*, (2020), la Encuesta de Autorregulación de Ronqui *et al.* (2024); RESUMeV León *et al.* (2015) y Técnica de Identificación de Significado y de oraciones de Esmer & Güneş (2019). La aplicación de estos instrumentos requiere el monitoreo activo y análisis consciente por parte del lector de sus propios procesos, de modo que se evalúan habilidades como identificación de errores y autoevaluación de la comprensión.

Los cinco instrumentos de este conjunto de artículos representan 14,2 % del corpus total. Comparten y manifiestan el papel central de la metacognición, no obstante, cada uno tiene características únicas, por ejemplo, REMICOM solicita que las niñas y niños supervisen su proceso de CL, a través de la detección de incongruencias en textos (Barreyro *et al.*, 2020). Por su parte, el cuestionario de autorregulación, profundiza en el monitoreo y planificación, debido a que combina una escala Likert con trazas en línea que verifican el comportamiento del lector durante la lectura. En el caso de RESUMeV, este orienta al estudiante a la reorganización y revisión mental de la información para poder plasmarla en un resumen. Finalmente, los instrumentos de Esmer & Güneş, (2019), promueven que las niñas y niños analicen oraciones que se encuentran alteradas, ya que previamente desarrollan la lectura de doce enunciados, lo que fomenta la autorreflexión en un nivel literal e inferencial. En virtud de lo mencionado, estos instrumentos no solamente miden a los alumnos, también supervisan cómo controlan y se hacen conscientes del proceso de la CL (tabla 6).

**Tabla 6.** Lista de instrumentos que evalúan metacognición y autorregulación.

Instrumento Autor(es)	Descripción / Ítems	Aplicación/ Nivel de primaria.	Formato	Proceso evaluado
Cuestionario de Autorregulación (Ronqui <i>et al.</i> , 2024)	13 ítems con escala Likert del 1-4 sobre previsión, seguimiento, evaluación.	Grupal, tras lectura. Para 6º grado.	Likert	Metacognición (autorregulación)
REMICOM (Barreyro <i>et al.</i> , 2020)	10 textos breves de 5 oraciones, 4 presentan contradicciones, 4 con pseudopalabras, 2 sin errores.	Individual o grupal, sin límite de tiempo. 3º y 4º grado de primaria.	Textos en los que hay que señalar errores	Monitoreo de CL/Metacognición
RESUMeV (León <i>et al.</i> , 2015)	Un texto narrativo (542 palabras)	Individual	Resumen escrito evaluado con lista de cotejo.	Comprensión integrativa y síntesis
Técnica de Identificación de Significado (MIT) (Esmer & Güneş, 2019)	12 ítems	Individual, sin límite de tiempo. 4º grado de primaria	Opción binaria	Comprensión literal e inferencial
Técnica de Verificación de Oraciones (SVT) (Esmer & Güneş, 2019)	16 ítems con base en la lectura de textos narrativos y expositivos: con cambios en el significado y distractores.	Individual sin límite de tiempo. 4º grado de primaria	Opción binaria	Detección de errores de interpretación

Fuente: Elaboración propia.

Adicionalmente, algunos instrumentos se interesan en evaluar las habilidades sintácticas y morfológicas, a través de tareas como la técnica de juicio de gramaticalidad visual (Schindler *et al.*, 2018), la tarea de comprensión y repetición de oraciones (Poulsen *et al.*, 2021), y PEHMM, enfocado en habilidades metamorfológicas (Vega *et al.*, 2017). Resulta interesante que estos cuatro instrumentos se evalúan componentes del lenguaje que sientan las bases de la CL, pero difieren en sus formatos y los grados a los que se dirigen.

Mientras Poulsen *et al.* (2021) y Schindler *et al.* (2018) exploran la sintaxis mediante interfaces digitales en 4° y 6° grado respectivamente, Vega *et al.* (2017) profundiza en la morfología por medio de una prueba en papel sin límite de tiempo, en 4° grado de primaria. Lo que refleja cómo diferentes componentes del lenguaje pueden analizarse individualmente para reconocer aquellos procesos que subyacen a la CL (tabla 7).

**Tabla 7.** Lista de instrumentos que evalúan habilidades morfológicas o sintaxis.

Instrumento Autor(es)	Descripción / Ítems	Aplicación/ Nivel de primaria.	Formato	Proceso evaluado
Juicio de gramaticalidad visual (Schindler <i>et al.</i> , 2018)	38 oraciones y 2 de práctica; varían en complejidad, se presionan botones con las opciones “correcta/incorrecta”.	Individual en computadora para 4° grado	Opción binaria (botones)	Habilidad sintáctica
PEHMM (Habilidades Metamorfológicas) (Vega <i>et al.</i> , 2017)	4 subpruebas (15 ítems c/u): derivación, contextualización, pseudopalabras, análisis raíz.	Individual para 4° grado.	Selección múltiple	Conciencia morfológica
Repetición de oraciones (Poulsen <i>et al.</i> , 2021)	30 oraciones incrustadas; repiten para activar sintaxis más allá de memoria.	Oral	Repetición	Conocimiento sintáctico / memoria
Tarea de comprensión de oraciones (Poulsen <i>et al.</i> , 2021)	30 oraciones complejas; opción múltiple; ritmo controlado (oral).	Auditiva	Opción múltiple	Conocimiento sintáctico

Fuente: Elaboración propia.

Por su parte, se hallaron dos instrumentos que relacionan evaluaciones orales con registros escritos o preguntas abiertas tras la lectura en voz alta, de modo que proporcionan una valoración integral de la CL y a la vez miden otros procesos como la fluidez y precisión en la lectura, en sus niveles literal e inferencial. Entre estas evaluaciones destacan EMLE-

TALE 2000 de Marín-Marín *et al.* (2023) y Moreno-Guerrero *et al.* (2023) y el Inventario informal de lectura (Papatga & Ersoy, 2016) (tabla 8).

**Tabla 8.** Lista de pruebas que evalúan CL en voz alta.

Instrumento Autor(es)	Descripción / Ítems	Aplicación/ Nivel de primaria.	Formato	Proceso evaluado
EMLE-TALE 2000 (Marín-Marín <i>et al.</i> , 2023; Moreno Guerrero, 2023)	4 subpruebas: el apartado de CL se mide con lectura en voz alta y después se realizan preguntas a nivel literal e inferencial.	Individual en 5º y 6º grado de primaria.	Prueba escrita	Fluidez, decodificación, comprensión.
Inventario informal de lectura (Papatga & Ersoy, 2016)	Lectura de textos en voz alta con registro de errores, después se realizan 10 preguntas abiertas.	Individual 4º grado.	Dos secciones. La primera con un texto y la segunda con preguntas Abiertas	Fluidez lectora, comprensión literal y crítica. Tres niveles de clasificación: independiente, instruccional y nivel ansiedad.

Fuente: Elaboración propia.

### Propiedades psicométricas de los instrumentos de CL.

Los estudios del presente corpus han documentado las propiedades psicométricas de instrumentos diseñados para evaluar la CL en educación primaria, es decir, que demuestran una confiabilidad de aceptable a alta y, diferentes formas de validez como la de contenido, constructo, convergente y predictiva.

Resulta importante mencionar, que los instrumentos con formato de opción múltiple y pasajes de textos expositivos y narrativos han registrado evidencias de confiabilidad y validez. El TCL Cadime *et al.* (2015), el CLIP Pascual *et al.* (2021), el ACL-3 Martínez-Álvarez *et al.* (2020) y el Test objetivo castellano de Quispe-Morales, (2022) presenta coeficientes alfa de Cronbach entre 0,70 y 0,92, además de validez mediante análisis factorial confirmatorio (AFC), ajuste al modelo Rasch o juicio de expertos. En este grupo también se ubica el WIAT-IIUK-T RCS McCartney *et al.* (2015), prueba estandarizada británica que ha mostrado ser sensible a intervenciones escolares, y el GraLeV aplicado por Seifert & Paleczek, (2022), con confiabilidad de  $r = 0,79$  y validez convergente.

Por su parte, las pruebas de formato *Cloze* y de corrección de oraciones, que evalúan CL desde un enfoque más procesual y jerárquico como, por ejemplo: la Técnica de Identificación de Significado (MIT) y la Técnica de Verificación de Oraciones (SVT) (Esmer & Güneş, 2019), ambas con coeficientes KR-20 superiores a 0,70. También destacan el HyTeC-Cloze (Moncada *et al.*, 2024), con alfa entre .79 y .93, y el instrumento digital de comprensión de oraciones tipo *Cloze* de Jungjohann *et al.* (2018), validado con análisis de unidimensionalidad. Asimismo, el instrumento TORCH (Hay & Stevenson, 2024) obtuvo una confiabilidad test-retest entre .91 y .93, lo que respalda su fiabilidad como medida estandarizada. Aunado al instrumento CLP (Barreyro *et al.*, 2020; Figueroa-Sepúlveda *et al.*, 2018) que presentó validez discriminante a través de un análisis de varianza y regresiones que permiten predecir la CL.

Con referencia a instrumentos que examinan metacognición en la CL, el REMICOM (Barreyro *et al.*, 2020) se centra en el monitoreo lector, con alfa de Cronbach de 0,74 y estabilidad temporal ( $r = 0,60$ ). Igualmente, el RESUMeV (León *et al.*, 2015), basado en que los estudiantes realicen resúmenes después de leer textos, mostró excelente consistencia interna ( $\alpha = 0,86$ ) y fiabilidad interjueces alta. De manera similar, el Cuestionario de Autorregulación (Ronqui *et al.*, 2024) manifiesta una confiabilidad moderada ( $\alpha = 0,64$ ), adecuada para estudios exploratorios. En este mismo eje, la Tarea de repetición de oraciones y la Tarea de comprensión de oraciones (Poulsen *et al.*, 2021) presentaron alfas de 0,86 y 0,80 respectivamente.

En el ámbito de pruebas estandarizadas con fuerte respaldo técnico digitalizado, destacan STAR Reading citado en Topping (2018), con confiabilidades superiores a 0,90 en tres formas (test-retest, mitades, genérica), y ELFE II en Visser *et al.* (2023), con alfas entre 0,80 y 0,90. Por su parte, la prueba LegiLexi (Nordström *et al.*, 2025) integra múltiples pruebas, con niveles de confiabilidad altos (ICC entre 0,90 y 0,98) y validación empírica con una muestra amplia ( $n = 11,791$ ), evalúa literalidad, inferencia y comprensión global.

Con relación a evaluaciones con enfoque diagnóstico y cognitivo para CL, el DCRCA (Li *et al.*, 2021) aplicó modelos de diagnóstico cognitivo (G-DINA) con validez empírica y alta precisión en clasificación de perfiles, y el PEHMM Vega Rodríguez *et al.* (2017), que evalúa habilidades metamorfológicas, cuenta con validez convergente ( $r = 0,66$  con CLP) y validez de contenido mediante juicio de expertos.

Finalmente, 28,5 % ( $n=8$ ) de los artículos emplean instrumentos altamente conocidos que no informan datos específicos que acrediten validez o confiabilidad. No obstante, se

manifiesta que ya han sido validados e incluso estandarizados, entre estos: CLP de Barreyro *et al.* (2020) y Figueroa Sepúlveda *et al.* (2018); Cuestionario de Autorregulación de Ronqui *et al.* (2024); ECOMPLEC-Pri de Míguez-Álvarez *et al.* (2022); EMLE-TALE 2000 Marín-Marín *et al.* (2023); PROLEC-R Míguez-Álvarez *et al.* (2022) Ramos-Carrillo & Moreno-Pérez (2023); WIAT-IIUK-T RCS de McCartney *et al.* (2015) y Woodcock-Johnson Tests de O'Rourke *et al.* (2016) han sido validados para entornos específicos, por lo que registran confiabilidad adecuada o validez a través de clasificación diagnóstica y coherencia con criterios externos (tabla 9).

**Tabla 9.** Lista de instrumentos con propiedades psicométricas.

Instrumento/ Autor(es) del artículo	Propiedad psicométrica
ACL-3 (Martínez-Álvarez <i>et al.</i> , 2020)	Confiabilidad $\alpha = 0,81$
CLIP (Test de Comprensión Lectora para Intervención en Primaria) (Pascual <i>et al.</i> , 2021)	Confiabilidad (KR-20): $\geq 0,84$ VC: 0,76 ; Habilidad = 0,62 VEI: KMO = 0,89; Bartlett $\chi^2(2142) = 14\ 644$ $p < 0,001$ EFA $\rightarrow$ 3 factores (31,32 % var.) AFC textos cortos: CFI= 0,895 ; RMSEA= 0,024 AFC textos largos: ajuste insatisfactorio AI (Análisis de Ítems): Dificultad $p = 0,15-0,91$ (mayoría 0,30-0,70) Discriminación $r_{pb} = 0,20-0,40$
CLP (Complejidad Progresiva) (Barreyro <i>et al.</i> , 2020; Figueroa Sepúlveda <i>et al.</i> , 2018).	NE Validada y estandarizada.
Cuestionario de Autorregulación (Ronqui <i>et al.</i> , 2024)	NE
CBM-Maze de oraciones (Jungjohann <i>et al.</i> , 2018)	Validez de constructo: Unidimensionalidad confirmada. RMSEA= 0,013
DCRCA (Li <i>et al.</i> , 2021)	$\alpha = 0,78$
ECOMPLEC-Pri Evaluación de la Comprensión Lectora para primaria (Míguez-Álvarez <i>et al.</i> , 2022)	NE Manifiestan validez.

ELFE II Test de Comprensión Lectora 2ª edición. (Visser <i>et al.</i> , 2023)	Confiabilidad Alfa de Cronbach entre 0,88 y 0,90.
EMLE-TALE 2000 (Marín-Marín <i>et al.</i> , 2023; Moreno 2023)	NE Manifiestan que es un instrumento validado.
GraLeV (Seifert & Paleczek, 2022)	Coefficiente de confiabilidad $r = 0,79$
HyTeC-Cloze en español. (Moncada <i>et al.</i> , 2024)	Alfa de Cronbach entre 0,79 y 0,93. Validez convergente: coeficiente de Spearman $\rho = 0,42$
Inventario Informal de Lectura (Papatga & Ersoy, 2016)	Coefficiente de correlación de $r = 0,979$ ( $p < .01$ ), lo que indica muy alta consistencia entre evaluadores.
Juicio de Gramaticalidad Visual (Schindler <i>et al.</i> , 2018)	RMSEA = 0,013 (IC 90 %: 0,009–0,016); CFI = 0,988; Gamma-hat = 0,991
LegiLexi (Nordström <i>et al.</i> , 2025)	Alta confiabilidad: ICC (consistencia intraclase) entre 0,90 y 0,98 Validez empírica en 11 791 estudiantes con diseño longitudinal.
PEHMM (Habilidades Metamorfológicas) (Vega Rodríguez <i>et al.</i> , 2017)	Validez de contenido: juicio de 8 expertos Validez de criterio: correlación con subprueba morfología de BLOC-S. Validez convergente: $r = ,66$ con prueba CLP.
PROLEC-R (Míguez-Álvarez <i>et al.</i> , 2022; Ramos-Carrillo & Moreno-Pérez, 2023)	Prueba estandarizada y validada en España.
REMICOM (Barreyro <i>et al.</i> , 2020)	Confiabilidad Alfa de Cronbach = 0,74 Kuder-Richardson-20 = 0,73 Estabilidad temporal (test-retest): $r = 0,60$ , $p < 0,001$
Repetición de Oraciones (Poulsen <i>et al.</i> , 2021)	Confiabilidad $\alpha = 0,86$ Análisis factorial confirmatorio (unidimensionalidad)
RESUMeV (León <i>et al.</i> , 2015)	Alfa de Cronbach total: 4.º primaria: 0,86 6.º primaria: 0,84 Total, general: 0,86 Fiabilidad interjueces (correlación de Pearson): De: 0,69 a 0,97

	Análisis de ítems: Medias de dificultad por criterio: 0,53 – 1,24 (según criterio y curso)
STAR Reading (Topping, 2018)	Confiabilidad: Fiabilidad test-retest: 0,92 Fiabilidad de división por mitades: 0,91 Fiabilidad genérica: 0,97 Prueba estandarizada de EE.UU.
Técnica de identificación de significado (MIT) (Esmer & Güneş, 2019)	Confiabilidad KR-20 = 0,77 para textos narrativos KR-20 = 0,73 para textos informativos Validez predictiva: R = 0,46; R <sup>2</sup> = 0,21, p < 0,05
Técnica de Verificación de Oraciones (SVT) (Esmer & Güneş, 2019)	Confiabilidad KR-20 = 0,74 para textos narrativos KR-20 = 0,78 para textos informativos Validez predictiva: R = 0,51; R <sup>2</sup> = 0,26, p < 0,05
Test Objetivo Castellano (Quispe-Morales, 2022)	Alfa de Cronbach: 0,921 Validez de contenido: juicio de 5 expertos 0,85
TCL Test de comprensión lectora (Cadime <i>et al.</i> , 2015)	Unidimensionalidad confirmada: CFI = 0,96; TLI = 0,96; RMSEA = 0,02 Ajuste al modelo Rasch (Infit y Outfit entre 0,5 y 1,5) PSR: 0,70 (TCL-2), 0,78 (TCL-3), 0,79 (TCL-4) ISR: 0,98 (en las tres formas).
TORCH (Hay & Stevenson, 2024)	Coefficiente de fiabilidad test-retest entre 0,91 y 0,93 Validez concurrente con prueba australiana.
WIAT-IIUK-T RCS (McCartney <i>et al.</i> , 2015)	NE Prueba estandarizada británica.
Woodcock-Johnson Tests (Batería breve) (O'Rourke <i>et al.</i> , 2016)	NE Manifiesta el uso de pruebas estandarizadas.

Fuente: Elaboración propia. Nota: EFA= Análisis factorial exploratorio. AFC= Análisis factorial confirmatorio. NE= No especificado. VC= Validez de contenido. VEI= Validez de estructura interna

## Discusión

A través del corpus analizado, se constata que existe una amplia variedad de instrumentos validados para evaluar la CL a nivel primaria, se pueden identificar cinco conjuntos de pruebas: con textos narrativos y expositivos, con espacios en blanco tipo *Cloze*, evaluaciones de metacognición y autorregulación, de sintaxis y metamorfológicas y de lectura en voz alta. Así también, se observa predominancia de métodos cuantitativos, ya que se centran en el análisis y la cuantificación de ítems de opción múltiple con estructuras jerárquicas que llevan procesos de complejidad gradual, para que finalmente se les otorgue un valor numérico. En términos generales, se comprobó que dichos artículos contaron con pruebas de mediana y alta confiabilidad y con diversas formas de validez que respaldaron su puesta en marcha para fines diagnósticos, de validación instrumental y como pruebas de apoyo para intervenciones con pre y postest.

Resulta interesante mencionar que a partir del 2018 se observa que las pruebas comenzaron a implementarse por vía digital, por lo que hay 35,7 % (n=10) de artículos que utilizan diversos software que facilitan las mediciones e interacciones (Jungjohann *et al.*, 2018; Li *et al.*, 2021; Míguez-Álvarez *et al.*, 2022; Moncada *et al.*, 2024; Nordström *et al.*, 2025; Poulsen *et al.*, 2021; Ronqui *et al.*, 2024; Seifert & Paleczek 2022; Schindler *et al.*, 2018; Visser *et al.*, (2023). Estas pruebas han señalado ventajas como el proveer resultados y retroalimentaciones instantáneas que disminuyen los errores manuales, al mismo tiempo que facilita el análisis de datos y trayectorias de los estudiantes.

Realizar un recuento de esta diversidad instrumental, resulta positivo para seleccionar de manera objetiva y contextualizada herramientas evaluativas según el nivel educativo, las necesidades grupales y de acuerdo con los fines de la evaluación. Como recomendación para docentes e investigadores se sugiere seleccionar pruebas con base en la evidencia psicométrica informada, su adecuado ajuste al contexto escolar, y la claridad con que operacionalizan los diversos niveles y procesos de la CL, para garantizar el derecho a una educación de calidad desde las primeras etapas escolares. Cuando se cuenta con recursos tecnológicos limitados, se sugiere optar por pruebas en formato de papel, pero de alta confiabilidad por ejemplo ACL-3, CLP, CLIP o PROLECR; en ambientes con plataformas, incorporar HyTeC-Cloze o DCRCA para obtener información diagnóstica inmediata.

## Conclusiones

El objetivo de la presente revisión sistemática se centró en el análisis de instrumentos validados para evaluar la CL a nivel primaria. A partir de la búsqueda en bases de datos, se hallaron 28 artículos con pruebas que cuentan con validez psicométrica y con evidencias de confiabilidad. Al realizar este mapeo, se obtuvo un panorama amplio sobre la diversidad de diseños instrumentales, su ubicación geográfica y distribución temporal, por lo que se observa que estos estudios se concentran en el continente europeo.

Entre los hallazgos más importantes se encuentra que predominan las pruebas que emplean textos narrativos con opción múltiple, las cuales son evaluaciones más tradicionales y empleadas a lo largo de décadas. Al mismo tiempo, se ha evidenciado que existen instrumentos digitales que dan retroalimentaciones inmediatas, cuestionarios y tareas orientadas al monitoreo de los lectores y al proceso de la autorregulación; lo que genera una mayor conciencia sobre la CL en los estudiantes

Por otra parte, si bien permanecen las implementaciones de evaluaciones clásicas, cada vez se van diseñando más herramientas que facilitan tanto la aplicación como el análisis de datos, que incluso permiten el seguimiento de los avances de las niñas y niños. De este modo, resalta como aportación de la revisión sistemática, esta tendencia de innovar en el diseño de instrumentos validados durante los últimos 10 años.

Al igual, se identifica que existe una brecha, ante la ausencia de pruebas que en una misma evaluación integren los procesos cognitivos y lingüísticos imprescindibles para la CL, ya que esto limita la medición de una manera holística y simultánea de los componentes de esta habilidad, pues se hallaron diseños que se enfocan en dimensiones aisladas. En consecuencia, para evaluar de manera más completa se requieren varias pruebas que de modo separado conllevan más tiempo y recursos.

En virtud de lo ya mencionado, esta revisión brinda apoyo a profesionales de la educación primaria a seleccionar de manera sistemática y analítica la o las herramientas que se adapten y contribuyan en mayor medida a las necesidades de las alumnas y alumnos de su entorno. Aunado a esto, se sugiere la incorporación de estudios longitudinales que determinen la sensibilidad al cambio de estas pruebas, a lo largo de distintos grados escolares y contextos socioculturales.

## **Contribución a futuras líneas de investigación**

Con base en los resultados del estudio se plantea la necesidad de que las investigaciones sobre CL cuenten con instrumentos validados o estandarizados, que reporten datos numéricos que avalen su objetividad y faciliten que los lectores empleen dichas pruebas e incluso repliquen las metodologías con la confianza de que cuentan con un instrumento de alta fiabilidad. Así como, la posibilidad de presentar las características de estas herramientas de maneras más detalladas y descriptivas.

Por otro lado, se sugiere el uso de pruebas que logren integrar varios procesos cognitivos y lingüísticos subyacentes de la CL. Entre estos, se encuentra la denominación rápida, las funciones ejecutivas, conciencia fonológica, fluidez y precisión lectora, las habilidades sintácticas y morfológicas, entre otras. Resulta importante mencionar, que estos procesos forman parte de las bases teóricas, sin embargo, no se presentan pruebas con esta clase de diseños.

### **Limitaciones**

Entre las limitaciones del estudio se encuentra que el análisis de los artículos se llevó a cabo en un período de diez años (2015 a 2025) y en los idiomas inglés y español. En consecuencia, cualquier investigación de otro año o idioma queda fuera de lo plasmado en este artículo.

Aunado a esto, la mayoría de los estudios son de origen europeo, ya que al seguir rigurosamente los criterios de inclusión y exclusión varios estudios latinoamericanos quedaron descartados. Sin embargo, es de reconocerse la importancia de continuar la búsqueda y análisis de estas regiones para este campo de estudio.

Por otro lado, los artículos presentaron datos y características limitadas de los instrumentos, de modo que existen aspectos de los diseños y de las propiedades psicométricas que no fueron posibles de contemplar.

### **Agradecimientos**

El presente estudio fue realizado con apoyo del financiamiento de la Secretaría de Ciencias, Humanidades, Tecnología e Innovación (SECIHTI) como parte de la beca núm. 4037255 para estudios de posgrado.

## Referencias

- Álvarez Porroa, C. del P., Asencio García, I. A., Chipana, H., & Lapoint Montes, V. A. (2022). Revisión sistemática acerca de la comprensión lectora en la etapa escolar. *Sinergias educativas*.  
<https://sinergiaseducativas.mx/index.php/revista/article/view/214/533>
- Austin, C. R., Wanzek, J., Scammacca, N. K., Vaughn, S., Gesel, S. A., Donegan, R. E., & Engemann, M. L. (2019). The relationship between study quality and the effects of supplemental reading interventions: A meta-analysis. *Exceptional Children*, 85(3), 347–366. <https://doi.org/10.1177/0014402918796164>
- Avendaño-Guevara, C., & Balderas-Gutiérrez, K. (2023). El contexto social y cultural en la lectoescritura de educación media superior. *Alteridad: Revista de educación*, 18(2), 39–54. <https://doi.org/10.17163/alt.v18n2.2023.03>
- Barreyro, J. P., Musci, M. C., Buonsanti, L., Brenlla, M. E., & Gottheil, B. (2020). REMICOM: evaluación del monitoreo de la comprensión lectora en niños y niñas. *Revista neuropsicología latinoamericana*, 12(2), 24–30.  
[https://www.neuropsicolatina.org/index.php/Neuropsicologia\\_Latinoamericana/article/view/592](https://www.neuropsicolatina.org/index.php/Neuropsicologia_Latinoamericana/article/view/592)
- Bettany-Saltikov, J. (2012). *How to do a Systematic Literature Review in Nursing A step-by-step guide*. Open University Press.
- Cabero Almenara, J., Piñero Virués, R., & Reyes Rebollo, M. M. (2018). Material educativo multimedia para el aumento de estrategias metacognitivas de comprensión lectora. *Perfiles educativos*, XL(159), 144–157.  
<https://doi.org/10.22201/iisue.24486167e.2018.159.58042>
- Cadime, I., Ribeiro, I., Viana, F. L., Santos, S., & Prieto, G. (2015). Calibración de un test de comprensión lectora para alumnos portugueses. *Anales de psicología*, 30(3), 1025–1034. <http://doi.org/10.6018/analesps.30.3.172611>
- Chávez-Delgado, M., González-Vergara, S., & Sepúlveda-López, F. (2022). Revisión sistemática de literatura sobre programas de intervención en habilidades de lectura inicial. *Páginas de educación*, 15(2), 98–127.  
<https://doi.org/10.22235/pe.v15i2.2775>
- Díaz Calle, Z., Noria Aliaga, V. M., & Buendía Molina, M. A. (2024). Comprensión lectora en la era digital: Una revisión sistemática. *Revista andina de educación*, 7(2), Artículo 000721. <https://doi.org/10.32719/26312816.2024.7.2.1>

- Esmer, B., & Güneş, A. M. (2019). El uso de la identificación de significado: Técnica para medir la lectura, habilidades de comprensión. *International Electronic Journal of Elementary Education*, 11(4), 413–420. <https://doi.org/10.26822/iejee.2019450836>
- Figueroa-Sepúlveda, S., & Gallego-Ortega, J. L. (2018). Vocabulario y comprensión lectora en escolares chilenos de primer ciclo básico. *Ocnos: Revista de estudios sobre lectura*, 17(1), 32–42. [https://doi.org/10.18239/ocnos\\_2018.17.1.1521](https://doi.org/10.18239/ocnos_2018.17.1.1521)
- Fuchs, L. S., Seethaler, P. M., Fuchs, D., & Espinas, D. (2023). Severe pandemic learning loss and the promise of remotely delivered intervention in students with comorbid reading and mathematics learning difficulty. *Journal of Learning Disabilities*, 56(4), 278–294. <https://doi.org/10.1177/00222194231170313>
- Gallegos de Dios, O. (2024). Analfabetismo funcional o barreras de aprendizaje en secundarias de México. *Revista diálogos e perspectivas em educação especial*, 11(1). <https://doi.org/10.36311/2358-8845.2024.v11n1.e0240012>
- Hay, I., & Stevenson, Y. M. (2024). El nivel de previsibilidad de los logros de lectura anteriores y la autopercepción y las medidas de comportamiento sobre el logro de lectura de los niños. *Australian Journal of Education.*, 68(2), 124–144. <https://doi.org/10.1177/00049441241253027>
- Hevia, F., Vergara-Lope, S., Velásquez-Durán, A., & Calderón, D. (2022). Estimation of the fundamental learning loss and learning poverty related to COVID-19 pandemic in Mexico. *International Journal of Educational Development*, 88, 102515. <https://doi.org/10.1016/j.ijedudev.2021.102515>
- Jungjohann, J., DeVries, J. M., Mühlhng, A., & Gebhardt, M. (2018). Using theory-based test construction to develop a new curriculum-based measurement for sentence reading comprehension. *Frontiers in Education*, 3, Article 115. <https://doi.org/10.3389/educ.2018.00115>
- Li, Y., Zhen, M., & Liu, J. (2021). Desarrollo y validación de la Evaluación Diagnóstica de Comprensión Lectora en Chino (DCRCA) para estudiantes de primaria bajo el marco de la evaluación diagnóstica cognitiva. *Frontiers in Psychology*, 12, 786612. <https://doi.org/10.3389/fpsyg.2021.786612>
- Lozada, G., Martínez, B., Acle, T., & Ordaz, V. (2016). Validez y confiabilidad de una prueba para evaluar el desempeño en lectura y escritura en niños de primaria. Maestría en Psicología, Residencia en Educación Especial, Universidad Nacional

- Autónoma de México (UNAM), Facultad de Estudios Superiores Zaragoza, Ciudad de México, México. <http://hdl.handle.net/10045/64226>
- Marín-Marín, J. A., López-Belmonte, J., Lampropoulos, G., & Moreno-Guerrero, A. J. (2023). Influencia de la aplicación de un plan lector sobre la motivación, la inteligencia emocional, la fluidez y la comprensión lectora en alumnos de Educación Primaria de España. *Texto libre*, 16, e41548. <https://doi.org/10.1590/1983-3652.2023.41548>
- Martínez-Álvarez, I., Llamas-Salguero, F., & López-Fernández, V. (2020). Relación entre aspectos neuropsicológicos cognitivos y socioemocionales en alumnos de Educación Primaria. *Revista Fuentes*, 22(2), 224–237. <https://doi.org/10.12795/revistafuentes.2020.v22.i2.08>
- McCartney, E., Boyle, J., & Ellis, S. (2015). Developing a universal reading comprehension intervention for mainstream primary schools within areas of social deprivation for children with and without language-learning impairment: A feasibility study. *International Journal of Language & Communication Disorders*, 50(1), 129–135. <https://doi.org/10.1111/1460-6984.12124>
- Míguez-Álvarez, C., Cuevas-Alonso, M., Saavedra, A., & Cabanach, R. G. (2022). El papel de las características del texto en la comprensión lectora de niños de primaria en español. *Revista iberoamericana de psicología y salud*, 13(1), 41–55. <https://doi.org/10.23923/j.rips.2022.01.053>
- Moncada, F., Ibáñez, R., Santana, A., & Guerra, C. (2024). Spanish adaptation of a cloze procedure to assess reading comprehension beyond the sentence level. *Reading and Writing*, 37(1), 165–186. <https://doi.org/10.1007/s11145-023-10488-2>
- Moreno, B., Muñoz, M., Cuellar, J., Domancic, S. & Villanueva, J. (2018). Revisiones Sistemáticas: definición y nociones básicas. *Revista clínica de periodoncia, implantología y rehabilitación oral*, 11(3), 184-186. <https://dx.doi.org/10.4067/S0719-01072018000300184>
- Moreno-Guerrero, A.-J., Marín-Marín, J.-A., Carmona-Serrano, N., & López-Belmonte, J. (2023). Impacto del Plan Lector en estudiantes de educación primaria. *Innoeduca. International Journal of Technology and Educational Innovation*, 9(2), 149–161. <https://doi.org/10.24310/innoeduca.2023.v9i2.15175>
- O'Rourke, D., Olshtroon, A., & O'Halloran, C. (2016). La Iniciativa de Lectura de Limerick: Una intervención de lectura dirigida a lectores con dificultades en la

escuela primaria. *Support for Learning*, 31(2), 150–162.

<https://doi.org/10.1111/1467-9604.12123>

Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura

(UNESCO). (2023). *Día internacional de la alfabetización. 8 de septiembre.*

<https://www.unesco.org/es/days/literacy>

Page, M., McKenzie, J., Bossuyt, P., Boutron, I., Hoffmann, T., Mulrow, C., Shamseer, L., Tetzlaff, J., Akl, E., Brennan, S., Chou, R., Glanville, J., Grimshaw, J., Hróbjartsson, A., Lalu, M., Li, T., Loder, E., Mayo-Wilson, E., McDonald, S., ... Alonso-Fernández, S. (2021). Declaración PRISMA 2020: una guía actualizada para la publicación de revisiones sistemáticas. *Revista española de cardiología*, 74(9), 790–799. <https://doi.org/10.1016/j.recesp.2021.06.016>

Papatga, E., & Ersoy, A. (2016). Improving reading comprehension skills through the Scratch program. *International Electronic Journal of Elementary Education*, 9(1), 124–150. <https://eric.ed.gov/?id=EJ1126664>

Pascual, G., Goikoetxea, E., & Bustos, H. (2021). Propiedades psicométricas de un Test de Comprensión Lectora para Intervenir en Primaria (CLIP). *Psykhé*, 30(1), 1–15. <https://doi.org/10.7764/psykhe.2018.22337>

Quispe-Morales, R. A. (2022). Desarrollo de la conciencia metalingüística y su efecto en la comprensión lectora en castellano como segunda lengua. *Revista electrónica educare*, 26(2), 1–19. <https://doi.org/10.15359/ree.26-2.11>

Ramírez Cubas, M. E., Condori Machaca, J. E., & Cañari Marticorena, H. F. (2025). Comprensión lectora en estudiantes de educación básica: Una revisión sistemática. Horizontes. *Revista de investigación en ciencias de la educación*, 9(número extraordinario), 1224–1238. <https://doi.org/10.33996/revistahorizontes.v9i37.980>

Ramos-Carrillo, J. M., & Moreno-Pérez, F. J. (2023). Detecting reading difficulties in Spanish in older elementary students in the context of the Response to Intervention model. *Dyslexia*, 29(4), 330–346. <https://doi.org/10.1002/dys.1755>

Ronqui Leites, V., Trías Seferian, D., & Huertas Martínez, J. A. (2024). Assessing self-regulated processes: What do primary school students do, say and think in the process of understanding a text? *European Journal of Psychology of Education*, 39, 2531–2550. <https://doi.org/10.1007/s10212-024-00830-9>

Schindler, J., Richter, T., Isberner, M.-B., Naumann, J., & Neeb, Y. (2018). Construct Validity of a Process-Oriented Test Assessing Syntactic Skills in German Primary

- School children. *Language Assessment Quarterly*, 15(2), 183–203.  
<https://doi.org/10.1080/15434303.2018.1446142>
- Seifert, S., & Paleczek, L. (2022). Comparing tablet and print mode of a German reading comprehension test in grade 3: Influence of test order, gender and language. *International Journal of Educational Research*, 113, Article 101948.  
<https://doi.org/10.1016/j.ijer.2022.101948>
- Snow, A. B., Morris, D., & Perney, J. (2018). Evaluating the effectiveness of a state-mandated benchmark reading assessment: mCLASS Reading 3D (Text Reading and Comprehension). *Reading Psychology*. 303-334  
<https://doi.org/10.1080/02702711.2017.1422302>
- Tamay Loaiza, N. (2024). *Impacto del rezago educativo en el proceso de enseñanza-aprendizaje* [Tesis de maestría, Universidad Politécnica Salesiana]. Universidad Politécnica Salesiana. <http://dspace.ups.edu.ec/handle/123456789/28081>
- Tennent, W. (2020). The assessment of reading comprehension in English primary schools: Investigating the validity of the Key Stage 2 reading standard assessment test (SAT). *Education 3-13*, 49. <https://doi.org/10.1080/03004279.2020.1735472>
- Thomas, D., Hopwood, B., Hatisaru, V., & Hicks, D. (2024). Diferencias de género en el rendimiento en lectura y aritmética a lo largo de los años escolares. *The Australian Educational Researcher*, 51(41-66). <https://doi.org/10.1007/s13384-022-00583-8>
- Topping, K. (2018). Implementation fidelity in computerised assessment of book reading. *Computers & Education*, 116, 176–190.  
<https://doi.org/10.1016/j.compedu.2017.09.009>
- UNICEF (2022). Cuatro de cada cinco niños y niñas en América Latina y el Caribe no podrán comprender un texto simple. Informe publicado por el Banco Mundial y UNICEF, en colaboración con la UNESCO. Ciudad de Panamá/Washington, D.C., <https://www.unicef.org/lac/comunicados-prensa/cuatro-de-cada-cinco-ninos-y-ninas-en-america-latina-y-el-caribe-no-podran-comprender-un-texto-simple#:~:text=CIUDAD%20DE%20PANAM%C3%81%2FWASHINGTON%2C%20D.C.,en%20colaboraci%C3%B3n%20con%20la%20UNESC>.
- UNICEF (2022). *El 70 % de los niños de 10 años se encuentran en situación de pobreza de aprendizajes y no pueden leer y comprender un texto simple* UNICEF.

<https://www.unicef.org/cuba/comunicados-prensa/el-70-de-los-ninos-de-10-annos-se-encuentran-en-situacion-de-pobreza>

Visser, L., Carstachau, F., von Goldammer, A., Brandenburg, J., Timmerman, M., Hasselhorn, M., & Mähler, C. (2023). Measurement invariance in relation to first language: An evaluation of German reading and spelling tests. *Applied Measurement in Education*, 36(2), 115–131.

<https://doi.org/10.1080/08957347.2023.2201701>

Rol de Contribución	Autor (es)
Conceptualización	Adela Aurora Cervantes Buenfil Pedro José Canto Herrera (igual)
Metodología	Adela Aurora Cervantes Buenfil Pedro José Canto Herrera (igual)
Software	N/A
Validación	Adela Aurora Cervantes Buenfil Pedro José Canto Herrera (igual)
Análisis Formal	Adela Aurora Cervantes Buenfil Pedro José Canto Herrera (igual)
Investigación	Adela Aurora Cervantes Buenfil
Recursos	Adela Aurora Cervantes Buenfil Pedro José canto Herrera (igual)
Curación de datos	Adela Aurora Cervantes Buenfil Pedro José Canto Herrera (igual)
Escritura - Preparación del borrador original	Adela Aurora Cervantes Buenfil Pedro José Canto Herrera (igual)
Escritura - Revisión y edición	Adela Aurora Cervantes Buenfil Pedro José Canto Herrera (igual)
Visualización	Adela Aurora Cervantes Buenfil Pedro José Canto Herrera (igual)
Supervisión	Pedro José Canto Herrera
Administración de Proyectos	Adela Aurora Cervantes Buenfil Pedro José canto Herrera (igual)
Adquisición de fondos	N/A