



**UNIVERSIDAD DE CONCEPCIÓN**  
**ESCUELA DE EDUCACIÓN**

**MEDICIÓN DEL PENSAMIENTO CRÍTICO EN ESTUDIANTES  
DE 7 A 10 AÑOS: DISEÑO Y VALIDACIÓN DE UN  
INSTRUMENTO EN CONTEXTO MATEMÁTICO**

Tesis presentada a la Escuela de Educación de la Universidad de Concepción para optar al título profesional de profesora en Educación General Básica con mención en Matemática y Ciencias Naturales

**Autoras** : Josselyn Noelia Águila Ulloa  
Bárbara Anaís Erices García  
Carolina Angélica Salgado Salgado  
**Profesor guía** : Dr. Sergio Orlando Morales Candia  
**Comisión evaluadora:** Mag. Johana Espinoza Rivas  
Mag. María Aurora Gutiérrez

Diciembre, 2025  
Los Ángeles, Chile

*Esta tesis fue financiada por ANID – Fondecyt de Iniciación en Investigación, folio N° 11250799*

© 2025, Josselyn Noelia Águila Ulloa, Bárbara Anaís Erices García y Carolina

Angélica Salgado Salgado

Se autoriza la reproducción total o parcial del documento, con fines académicos, por cualquier medio o procedimiento, incluyendo la cita bibliográfica del documento.

## ÍNDICE

<b>I. INTRODUCCIÓN .....</b>	<b>10</b>
1.1 ANTECEDENTES DEL ESTUDIO .....	10
1.1.1 Propósitos del pensamiento crítico en la formación integral .....	11
1.1.2 Importancia del pensamiento crítico en las matemáticas .....	12
1.1.3 Factores influyentes en el desarrollo del pensamiento crítico .....	12
1.1.4 Evaluación del pensamiento crítico en educación primaria: Desafíos y limitaciones actuales .....	15
1.2 PLANTEAMIENTO DEL PROBLEMA .....	21
1.2.1 Justificación del estudio .....	23
1.3 PREGUNTAS Y OBJETIVOS DE INVESTIGACIÓN .....	24
1.3.1 Preguntas de investigación .....	24
1.3.2 Objetivos de investigación .....	25
1.3.3 Naturaleza del estudio .....	25
<b>II. MARCO CONCEPTUAL .....</b>	<b>27</b>
2.1 DEFINICIÓN DEL PENSAMIENTO CRÍTICO .....	27
2.2 COMPONENTES DEL PENSAMIENTO CRÍTICO .....	28
2.2.1 Habilidades del pensamiento crítico .....	28
2.2.2 Sub-habilidades del pensamiento crítico .....	31
2.2.3 Disposiciones al pensamiento crítico .....	32
2.3 CONTEXTUALIZACIÓN DE LA INFERENCIA ESTADÍSTICA .....	34
2.3.1 La necesidad del enfoque informal .....	35
2.3.2 Definición del razonamiento inferencial informal (RII) .....	35
2.3.3 El RII como un proceso de razonamiento .....	36
<b>III. METODOLOGÍA .....</b>	<b>38</b>
3.1 ENFOQUE Y DISEÑO DE LA INVESTIGACIÓN .....	38
3.2 PARTICIPANTES .....	39
3.3 DISEÑO Y CARACTERÍSTICAS DEL INSTRUMENTO .....	42
3.4 PROCEDIMIENTO DE VALIDACIÓN .....	42
3.4.1 Diseño de la escala .....	44
3.4.2 Validación de contenido .....	45

3.5 CONSIDERACIONES ÉTICAS .....	61
<b>IV. ANÁLISIS Y RESULTADOS .....</b>	<b>62</b>
4.1 RESULTADOS DERIVADOS DE LAS ENTREVISTAS COGNITIVAS .....	62
4.1.1 Comprensión semántica y lingüística.....	63
4.1.2 Comprensión de la tarea y estructura.....	68
4.1.3 Comprensión conceptual.....	69
4.1.4 Engagement y pertinencia percibida.....	71
4.2 RESULTADOS DERIVADOS DE LA EVALUACIÓN DE JUECES EXPERTOS .....	72
4.2.1 Síntesis por criterio .....	77
4.2.2 Análisis y ajustes realizados por ítem.....	85
4.2.3 Resultado de la validez de contenido según el coeficiente V de Aiken.....	113
<b>V. DISCUSIÓN.....</b>	<b>116</b>
5.1 Validez del instrumento en el contexto de Educación Primaria.....	116
<b>VI. CONCLUSIONES, LIMITACIONES Y PROYECCIONES DEL ESTUDIO</b>	<b>119</b>
6.1 Conclusiones sobre los objetivos y las preguntas de investigación .....	119
6.1.1 Conclusiones sobre el objetivo general .....	119
6.1.2 Conclusiones sobre el objetivo específico 1 .....	120
6.1.3 Conclusiones sobre el objetivo específico 2 .....	121
6.1.4 Conclusiones sobre el objetivo específico 3 .....	122
6.2 Limitaciones y proyecciones del estudio.....	123
6.2.1 Limitaciones del estudio.....	123
6.2.2 Proyecciones del estudio.....	124
<b>REFERENCIAS .....</b>	<b>126</b>
<b>BIBLIOGRAFÍA.....</b>	<b>133</b>
<b>ANEXOS .....</b>	<b>134</b>
8.1 ANEXO 1: INSTRUMENTO DE VALIDACIÓN .....	134
8.2 ANEXO 2: INSTRUMENTO PARA LA EVALUACIÓN DE ENTREVISTA COGNITIVA....	159
8.3 ANEXO 3: MANUAL DE PROTOCOLO PARA LA APLICACIÓN DE LA ENTREVISTA COGNITIVA.....	162

## ÍNDICE DE TABLAS

<b>Tabla 1</b> <i>Principales instrumentos internacionales de evaluación del pensamiento crítico</i> .....	15
<b>Tabla 2</b> <i>Habilidades y sub-habilidades del pensamiento crítico de acuerdo con el modelo de Facione (1990; 2020)</i> .....	31
<b>Tabla 3</b> <i>Síntesis conceptual del pensamiento crítico</i> .....	47
<b>Tabla 4</b> <i>Ejemplo del criterio de “Claridad” presentado en “Instrumento de validación”</i> .....	49
<b>Tabla 5</b> <i>Criterios de interpretación del coeficiente V de Aiken.</i> .....	53
<b>Tabla 6</b> <i>Rúbrica de criterios para la evaluación de la entrevista cognitiva</i> .....	54
<b>Tabla 7</b> <i>Escala Likert para la evaluación del grado de adecuación de los ítems del instrumento</i> .....	56
<b>Tabla 8</b> <i>Valoraciones de entrevistas cognitivas según criterios de evaluación</i> .....	62
<b>Tabla 9</b> <i>Puntuaciones obtenidas en ítems críticos en la primera evaluación de jueces expertos</i> .....	73
<b>Tabla 10</b> <i>Puntuaciones obtenidas en ítems críticos en la segunda evaluación de jueces expertos</i> .....	76
<b>Tabla 11</b> <i>Puntuaciones obtenidas de ítems críticos y en revisión para el criterio de “Claridad” en la primera y segunda evaluación de jueces expertos</i> .....	78
<b>Tabla 12</b> <i>Puntuaciones obtenidas de ítems críticos y en revisión para el criterio de “Pertinencia” en la primera y segunda evaluación de jueces expertos</i> .....	79
<b>Tabla 13</b> <i>Puntuaciones obtenidas de ítems críticos y en revisión para el criterio de “Contextualización para el razonamiento inferencial informal” en la primera y segunda evaluación de jueces expertos</i> .....	81
<b>Tabla 14</b> <i>Puntuaciones obtenidas en ítems críticos y en revisión para el criterio de “Relevancia de las alternativas” en la primera y segunda evaluación de jueces expertos</i> .....	83

<b>Tabla 15</b> <i>Puntuaciones obtenidas en ítems críticos y en revisión para el criterio de “Cobertura del pensamiento crítico” en la primera y segunda evaluación de jueces expertos</i> .....	84
<b>Tabla 16</b> <i>Resultados del coeficiente V de Aiken</i> .....	114

## ÍNDICE DE ILUSTRACIONES

<b>Figura 1</b> <i>Esquema sobre la metodología de la investigación</i> .....	38
<b>Figura 2</b> <i>Esquema metodológico sobre los participantes</i> .....	39
<b>Figura 3</b> <i>Esquema sobre el proceso de validación del instrumento de medición</i> .....	42
<b>Figura 4</b> <i>Esquema del diseño de la escala del instrumento de medición</i> .....	44
<b>Figura 5</b> <i>Esquema sobre el proceso de validación de contenido del instrumento de medición</i> .....	46
<b>Figura 6</b> <i>Ejemplo de ficha de evaluación por ítem</i> .....	50
<b>Figura 7</b> <i>Preguntas de sondeo protocolarias por criterio</i> .....	58
<b>Figura 8</b> <i>Ejemplo de hoja de cálculo con datos cuantitativos del ítem 1 y 2</i> .....	60
<b>Figura 9</b> <i>Versión original del ítem 3 utilizado en el “Instrumento para la evaluación de Entrevista Cognitiva”</i> .....	64
<b>Figura 10</b> <i>Versión original del ítem 4 utilizado en el instrumento “Instrumento para la evaluación de Entrevista Cognitiva”</i> .....	65
<b>Figura 11</b> <i>Versión original del ítem 9 utilizando en el “Instrumento para la evaluación de Entrevista Cognitiva”</i> .....	66
<b>Figura 12</b> <i>Versión original del ítem 13 utilizando en el “Instrumento para la evaluación de Entrevista Cognitiva”</i> .....	67
<b>Figura 13</b> <i>Comparación entre el ítem 5 original y su modificación</i> .....	91
<b>Figura 14</b> <i>Comparación entre el ítem 11 original y su modificación</i> .....	100
<b>Figura 15</b> <i>Comparación entre el ítem 13 original y su modificación</i> .....	104
<b>Figura 16</b> <i>Comparación entre el ítem 15 original y su modificación</i> .....	108
<b>Figura 17</b> <i>Comparación entre el ítem 16 original y su modificación</i> .....	111

## RESUMEN

El estudio tuvo como propósito diseñar y validar un instrumento para medir el pensamiento crítico en estudiantes de segundo a cuarto año básico (7 a 10 años) en el contexto del razonamiento inferencial informal en matemáticas. Frente a la escasa disponibilidad de instrumentos válidos para la evaluación de esta competencia en educación primaria, se desarrolló una investigación de enfoque mixto, con diseño no experimental transeccional descriptivo, orientada a la elaboración y validación de contenido. El instrumento creado corresponde a un test de 16 ítems de opción múltiple, construido a partir del modelo de pensamiento crítico de Facione y alineado con el currículum nacional de Matemáticas. La validación de contenido se realizó mediante juicio de expertos y entrevistas cognitivas, con la participación de siete jueces expertos en psicología educativa y didáctica de la matemática y diez estudiantes pertenecientes al sistema educacional chileno. Se calculó el coeficiente V de Aiken, obteniendo valores entre 0,819 y 0,943, todos iguales o superiores al punto de corte para considerarse adecuados ( $V \geq 0,80$ ). Se concluye que el instrumento presenta evidencia sólida de validez de contenido para evaluar el pensamiento crítico matemático en estudiantes chilenos de 7 a 10 años.

## PALABRAS CLAVES

pensamiento crítico, validación de instrumentos, matemáticas, educación primaria, razonamiento inferencial informal.

## ABSTRACT

The study aimed to design and validate an instrument to measure critical thinking in students from second to fourth grade (7 to 10 years old) within the context of informal inferential reasoning in mathematics. Faced with the scarce availability of valid instruments to assess this competence in primary education, a mixed-methods investigation was conducted, using a non-experimental, cross-sectional descriptive design focused on test development and content validation. The instrument created is a 16-item multiple-choice test, built on Facione's critical thinking model and aligned with the national mathematics curriculum. Content validation was carried out through expert judgment and cognitive interviews, involving seven experts in educational psychology and mathematics education and ten students from the Chilean school system. Aiken's V coefficient was calculated, yielding values between 0.819 and 0.943, all equal to or above the cut-off value for adequacy ( $V \geq 0.80$ ). The study concludes that the instrument provides solid evidence of content validity for assessing mathematical critical thinking in Chilean students aged 7 to 10 years.

## KEY WORDS

critical thinking, instrument validation, mathematics, primary education, informal inferential reasoning.

## I. INTRODUCCIÓN

### 1.1 ANTECEDENTES DEL ESTUDIO

En la actualidad, el pensamiento crítico es reconocido como una competencia esencial para enfrentar los desafíos del siglo XXI. La UNESCO (2025) lo sitúa junto a habilidades como la creatividad, la innovación y la resiliencia, necesarias para adaptarse a un mundo en constante transformación social, económica y tecnológica. Esta importancia ha llevado a instituciones educativas a incorporarlo como un pilar formativo (Deng et al., 2023), dado su papel crucial en el desarrollo de capacidades de razonamiento y es un elemento clave para desenvolverse en una sociedad moderna y compleja (Hang, 2020).

A pesar de su relevancia, su enseñanza aún se concentra en la educación superior, descuidando su integración sistemática en niveles primarios (Deng et al., 2023). Esto contrasta con la evidencia que resalta la necesidad de cultivarlo desde edades tempranas, no solo por su vinculación con competencias como la comprensión lectora (Lin et al., 2017), sino también porque prepara a los estudiantes para navegar en entornos globalizados y tecnológicamente dinámicos (Wijayanti et al., 2018).

Esta perspectiva es compartida por marcos normativos, como las Bases Curriculares de Primero a Sexto Básico (Ministerio de Educación de Chile [MINEDUC], 2024), que promueven aprendizajes con el propósito de adquirir autonomía para participar en la sociedad, desarrollando pensamiento crítico y creativo, junto con capacidades de comunicación y reflexión. Además, enfatizan la necesidad de enseñar a los estudiantes a buscar, seleccionar, evaluar y estructurar información, aplicando un pensamiento crítico y una actitud analítica frente a la sobreabundancia de datos, habilidades clave en una era de sobreabundancia de datos.

### 1.1.1 Propósitos del pensamiento crítico en la formación integral

El pensamiento crítico se define como un proceso intencional dirigido a fines específicos, como resolver problemas, interpretar significados o evaluar argumentos (Facione, 2020). Este tiene como objetivo último formar aprendices permanentes y ciudadanos capaces, lo que evidencia su importancia crucial más allá del ámbito académico, para la vida personal y social. (Hang, 2020).

En este marco, el pensamiento crítico influye en dimensiones esenciales del desarrollo personal: el espíritu crítico, que fomenta la curiosidad y la capacidad de formular preguntas y buscar información confiable (Facione, 2020); la independencia intelectual, que promueve el aprendizaje autónomo y la generación de ideas propias, sin depender de la enseñanza tradicional de terceros (Choque, 2019); y el juicio razonado, que impulsa a emitir opiniones fundamentadas en buenas razones, evitando suposiciones o creencias infundadas (Nomen, 2019).

Su desarrollo no sólo enriquece el razonamiento individual, sino que también se correlaciona con un mejor desempeño académico. Deng et al. (2023) evidencian que estudiantes con habilidades críticas más sólidas adquieren conocimientos con mayor eficacia, lo que sustenta su éxito educativo. Además, como señala Choque (2019), el pensamiento crítico fomenta la independencia intelectual, es decir, la capacidad de cuestionar, analizar y tomar decisiones conscientes, tanto en contextos de aprendizaje como en la sociedad.

Más allá de lo individual, el pensamiento crítico tiene un propósito colectivo. Según Nomen (2019), su meta no se limita a resolver problemas técnicos, sino a promover sociedades más justas y libres mediante el diálogo, la cooperación y el respeto a la diversidad. Facione (2020) va más allá, describiéndolo como un motor de progreso humano, esencial en la transición desde formas de pensamiento rudimentarias hacia una sensibilidad global. Así, su valor culmina en la capacidad de inspirar cambios sociales basados en la equidad y la reflexión crítica.

### 1.1.2 Importancia del pensamiento crítico en las matemáticas

En el ámbito específico de las matemáticas, el pensamiento crítico trasciende la mera aplicación de algoritmos, fomentando habilidades como el análisis, la justificación y la toma de decisiones fundamentadas. Según Sosa-Gutiérrez et al. (2023), esta competencia permite a los estudiantes resolver problemas de manera lógica, comprender conceptos con profundidad y aplicar conocimientos en contextos reales, adoptando una actitud activa y reflexiva.

No obstante, como advierten Gallegos et al. (2024), la enseñanza tradicional en primaria suele centrarse en procedimientos memorísticos, limitando la capacidad de los alumnos para transferir sus conocimientos a situaciones prácticas. Para superar esto, proponen problemas contextualizados que conecten las matemáticas con la vida cotidiana, potenciando así el razonamiento crítico.

La eficacia de este enfoque se sustenta en habilidades como la flexibilidad cognitiva, definida por Güner y Gökçe (2021) como la capacidad de adaptar el conocimiento a nuevas demandas, un factor crucial para el éxito matemático. Esta habilidad permite a los estudiantes reinterpretar problemas, explorar múltiples soluciones y ajustar estrategias, consolidando un aprendizaje significativo y adaptable.

### 1.1.3 Factores influyentes en el desarrollo del pensamiento crítico

El desarrollo del pensamiento crítico en matemáticas depende de factores familiares, cognitivos y pedagógicos. Algunos favorecen su progreso, como el apoyo familiar, la flexibilidad cognitiva y metodologías activas; otros lo dificultan, como la sobreprotección parental, la ansiedad matemática y la enseñanza tradicional de baja demanda cognitiva. A continuación, se presentan estos factores positivos y negativos.

## A. Positivos:

- **Factores familiares positivos:** El contexto familiar juega un papel dual. Por un lado, el estatus socioeconómico, la educación parental y los estilos de crianza influyen significativamente. Deng et al. (2023) evidencian que hijos de padres con educación universitaria desarrollan mayores habilidades y disposiciones críticas, vinculado a acceso a recursos cognitivos y modelos de razonamiento.
- **La base cognitiva. Rendimiento académico y flexibilidad:** El rendimiento académico temprano es un predictor clave, ya que proporciona los conocimientos base para habilidades críticas (Deng et al., 2023). Además, la flexibilidad cognitiva emerge como un componente esencial para el pensamiento crítico matemático, permitiendo reinterpretar problemas y explorar soluciones múltiples (Güner y Gökçe, 2021).
- **Facilitadores pedagógicos. Metodologías activas y preguntas reflexivas:** Frente a estos desafíos, enfoques como el Aprendizaje Basado en Problemas (ABP) demuestran eficacia. De acuerdo con Wijayanti et al. (2018), la aplicación del ABP incrementa significativamente la participación activa de los estudiantes, promueve la capacidad de resolver problemas de manera autónoma y favorece una interacción más dinámica entre el docente y los alumnos. Este modelo logra así trascender las limitaciones de la enseñanza tradicional, transformando a los estudiantes de receptores pasivos en agentes activos de su propio aprendizaje. Asimismo, las prácticas docentes son determinantes:
  - **Preguntas de alta calidad:** Interrogantes que exigen comparación, interpretación o reflexión ética (Nomen, 2019), en lugar de respuestas cerradas.
  - **Enfoques flexibles:** Evitar encasillar a los estudiantes en patrones únicos (Güner y Gökçe, 2021) para promover flexibilidad cognitiva.

## B. Negativos:

- **Factores familiares negativos:** La sobreprotección parental se correlaciona negativamente con el pensamiento crítico, al restringir la autonomía y la disposición a enfrentar desafíos intelectuales (Deng et al., 2023). Estos hallazgos subrayan la necesidad de equilibrar apoyo y promoción de independencia.
- **Barreras afectivas. La ansiedad matemática:** Entre los factores que inhiben el pensamiento crítico, la ansiedad matemática, definida como miedo a situaciones que involucran esta disciplina (Ramírez et al., 2018), destaca por su impacto negativo. Estudios como los de Imbo y Vandierendonck (2007) muestran que estudiantes con alta ansiedad evitan estrategias complejas, optando por enfoques básicos que limitan el análisis y la justificación de soluciones. Esta ansiedad no solo reduce el rendimiento académico (Güner y Gökçe, 2021), sino que también debilita la relación entre flexibilidad cognitiva y pensamiento crítico, especialmente durante transiciones educativas. Además, se manifiesta en síntomas cognitivos (dificultad para concentrarse) y respuestas fisiológicas (Jiang et al., 2021), creando un círculo vicioso que obstaculiza el desarrollo de habilidades críticas.
- **Obstáculos pedagógicos. Enseñanza tradicional y baja demanda cognitiva:** La persistencia de métodos tradicionales es una barrera clave. Choque (2019) y Sparapani et al. (2023) coinciden en que muchos sistemas educativos priorizan la memorización y la aplicación mecánica de procedimientos, limitando oportunidades para el análisis o la evaluación de alternativas. A esto se suma:
  - **Prácticas pasivas:** Estudiantes con escasa iniciativa y participación (Wijayanti, 2018).
  - **Tareas de baja demanda cognitiva:** Actividades que no requieren razonamiento complejo (Sparapani et al., 2023).
  - **Dificultad de transferencia:** Habilidades críticas desarrolladas en un dominio (ej. matemáticas) no siempre se aplican a otros (Willingham, 2008).

#### 1.1.4 Evaluación del pensamiento crítico en educación primaria: Desafíos y limitaciones actuales

##### A. Problemas conceptuales y falta de instrumentos adecuados

El pensamiento crítico, aunque reconocido como competencia clave en educación, enfrenta importantes retos de evaluación debido a la ausencia de consenso teórico y herramientas específicas. La multiplicidad de definiciones y marcos conceptuales ha generado estándares dispares para su medición, particularmente en el dominio matemático (Evendi et al., 2023). Esta falta de claridad dificulta el desarrollo de instrumentos válidos y confiables para primaria, donde la evaluación de estas habilidades resulta crucial. En la Tabla 1 se describen los diversos instrumentos que han intentado abordar esta necesidad desde perspectivas generales del pensamiento crítico.

**Tabla 1**

*Principales instrumentos internacionales de evaluación del pensamiento crítico*

Watson-Glaser Critical Thinking Appraisal (WGCTA) (Bernard et al., 2008; Watson y Glaser, 1980)	
Autores/ Año	Goodwin Watson y Robert Glaser; 1920 publicación original, 2011 última edición.
Formato ítems/ Duración	Opción múltiple. 40 ítems, 45 minutos de duración.
Población	Universitarios.
Habilidades evaluadas	Inferir conclusiones, reconocer supuestos, deducir, interpretar, evaluar argumentos
Comentarios/ Limitaciones	Funciona mejor como medida global; no adecuado para primaria o el área de matemáticas.
Cornell Critical Thinking Tests (X-Level y Z-Level) (The Critical Thinking Co., 2005)	
Autores/ Año	Robert H. Ennis, Jason Millman y Thomas Norbert Tomko; 1985 publicación original, 2005 edición final.

Formato ítems/ Duración	Opción múltiple. 50 minutos de duración. Nivel X: 71 ítems; Nivel Z: 52 ítems.
Población	Nivel X: 10–18 años (Grados 5–12+); Nivel Z: 16+ años, universitarios y adultos
Habilidades evaluadas	Nivel X: inducción, deducción, credibilidad, supuestos; Nivel Z: semántica, definición, predicción
Comentarios/ Limitaciones	Funciona mejor como medida global; Nivel Z más complejo.

---

California Critical Thinking Disposition Inventory (CCTDI) (Insight Assessment, 2007)

---

Autores/ Año	Peter A. Facione y Noreen C. Facione; 1992 publicación original, 2007 edición final.
Formato ítems/ Duración	Likert. 75 ítems, 30 minutos de duración.
Población	Universitarios y adultos.
Habilidades evaluadas	Disposiciones: búsqueda de la verdad, mente abierta, analiticidad, sistematicidad, confianza en el razonamiento, curiosidad, madurez cognitiva.
Comentarios/ Limitaciones	Mide disposiciones al pensamiento crítico, no habilidades.

---

California Critical Thinking Skills Test (CCTST) (Insight Assessment, 2010)

---

Autores/ Año	Peter A. Facione; 1990 publicación original, 2010 edición final.
Formato ítems/ Duración	Opción múltiple. 40 ítems, 55 minutos de duración.
Población	Universitarios y adultos.
Habilidades evaluadas	Análisis, interpretación, inferencia, evaluación, explicación, inducción, deducción, alfabetización numérica.
Comentarios/ Limitaciones	Evalúa habilidades cognitivas; orientado a educación superior.

---

---

K-12 Reasoning & Imagination Scales for Thinking and Learning (Insight Assessment, s.f)

---

Autores/ Año	Peter A. Facione y Noreen C. Facione; no se crearon en un año específico, sino que forman parte de la serie de evaluaciones EDUCATE INSIGHT de la compañía.
Formato ítems/ Duración	Likert + opción múltiple. El número de ítems y el tiempo de duración son variables según el nivel.
Población	K-2 (entre 5 y 8 años), 3-5 (entre 8 y 11 años), 6-8 (entre 11 y 14 años) y 9-12 (entre 14 y 18 años).
Habilidades evaluadas	Pensamiento crítico y creativo, mentalidad de aprendizaje.
Comentarios/ Limitaciones	Adaptado al nivel cognitivo; útil para planificación curricular.

---

Halpern Critical Thinking Assessment (HCTA) (De Bie et al., 2015)

---

Autores/ Año	Dianne F. Halpern; 2006 publicación original validada, 2010 última edición.
Formato ítems/ Duración	Ensayo + opción múltiple. 25 escenarios, con una duración de 45 a 80 minutos.
Población	Universitarios y adultos.
Habilidades evaluadas	Razonamiento verbal, análisis de argumentos, pensamiento como prueba de hipótesis, probabilidad e incertidumbre, toma de decisiones y resolución de problemas.
Comentarios/ Limitaciones	Mejor como medida global; combina componente cognitivo y disposicional.

---

Ennis-Weir Critical Thinking Essay Test (Robert H. Ennis y Eric Weir, 1985)

---

Autores/ Año	Robert H. Ennis y Eric Weir; 1985 publicación original, 2005 última edición.
Formato ítems/ Duración	Ensayo. 1 ejercicio con una duración de 40 minutos.
Población	Secundaria y universitarios.

Habilidades evaluadas	Evaluación de argumentos complejos, identificar razones y suposiciones, ofrecer buenas razones, detectar falacias, ver otras posibilidades.
Comentarios/ Limitaciones	Holística; no evalúa pensamiento deductivo; no adecuado para primaria.

---

### **B. Brecha investigativa: Enfoque en educación superior vs. primaria**

En la revisión de Abrami et al. (2015), que reunió 341 mediciones de resultados de programas diseñados para desarrollar el pensamiento crítico, se observa una marcada diferencia en la cantidad de investigaciones realizadas según nivel educativo. Del total de mediciones analizadas, el 36,9% corresponde a estudios efectuados en educación universitaria de pregrado, mientras que sólo un 14,3% proviene de educación primaria. El resto se reparte entre secundaria baja (22,9%), secundaria alta (20,8%) y posgrado o adultos (5,0%). Es importante destacar que estos porcentajes reflejan la distribución de los estudios incluidos en la revisión, no la magnitud de las mejoras obtenidas, lo que evidencia que, aunque las intervenciones para fortalecer el pensamiento crítico pueden ser efectivas en todos los niveles, la investigación se concentra principalmente en la educación superior, dejando los primeros niveles de escolaridad considerablemente menos explorado.

Autores como Ossa-Cornejo et al. (2017) señalan que “a pesar de la variada existencia de instrumentos, la mayoría han sido utilizados en estudiantes universitarios y adultos (Pencrisal, HCTAES, TPC, CCTT, CTDI, CCTST) y unos pocos a estudiantes de secundaria, como el HCTAES” (p. 26). Esta idea se ve reforzada por Deng et al. (2023), quien indica que “cultivar el PC de los estudiantes es considerado un objetivo crucial de la educación contemporánea, con la mayoría de los esfuerzos centrados en la educación superior, y menos atención dada al PC en la educación primaria” (p. 11). En conjunto, ambas referencias evidencian que el pensamiento crítico ha sido abordado con mayor énfasis en la educación superior que en la primaria.

Una posible explicación para esta disparidad se encuentra en una concepción tradicional, influenciada por la teoría piagetiana, que consideraba a los niños pequeños cognitivamente inmaduros para las operaciones formales requeridas por el pensamiento crítico, asociándolo así a etapas más avanzadas del desarrollo (Massa, 2014).

### **C. Limitaciones metodológicas en la evaluación**

El informe de Lai (2011) advierte sobre las restricciones de los instrumentos disponibles para evaluar el pensamiento crítico en educación primaria. Señala que las pruebas estandarizadas más utilizadas, como el California Critical Thinking Skills Test, no fueron diseñadas para aplicarse antes de cuarto grado, evidenciando un vacío metodológico en los niveles iniciales. Además, cuestiona su validez y confiabilidad, dado que los resultados suelen depender más de las características específicas de las tareas que de una evaluación auténtica de la habilidad.

Por su parte, Ennis (2018), destacado teórico en el campo, manifiesta un marcado escepticismo respecto a la confiabilidad de los test de pensamiento crítico disponibles. Advierte que la evidencia sobre sus propiedades psicométricas es limitada y que muchos estudios de validación presentan deficiencias metodológicas, como la ausencia de grupos de control adecuados, lo que introduce variables alternativas que pueden distorsionar significativamente los resultados.

En conjunto, el análisis de Lai (2011), reforzado por las observaciones críticas de Ennis (2018), indica que actualmente el campo carece de herramientas sólidas y validadas para la etapa de primaria, lo que restringe considerablemente la capacidad de medir con rigor el desarrollo de esta competencia fundamental.

### **D. Consecuencias en la práctica educativa**

La falta de instrumentos adecuados para medir el pensamiento crítico, especialmente en áreas como matemáticas, representa un desafío importante (Evendi et al., 2023). Esta carencia genera tres problemas principales:

- **Evaluaciones imprecisas:** La ambigüedad en los instrumentos provoca que sea imposible distinguir si los profesores están midiendo el éxito en las habilidades de pensamiento crítico o solo las habilidades cognitivas normales generales (Evendi et al., 2023).
- **Dificultades en la enseñanza:** Sin una evaluación clara, es difícil conocer el impacto de la instrucción, y los objetivos de pensamiento crítico vagos o implícitos tienen poco efecto en el desarrollo de estas habilidades (Abrami et al., 2008).
- **Percepción de la edad:** Se mantiene la idea de que el pensamiento crítico es solo para alumnos mayores, debido a su asociación histórica con jóvenes y adultos (Massa, 2014).

Esta situación genera un círculo vicioso que limita la mejora educativa. Por ello, resulta fundamental desarrollar y validar instrumentos específicos de evaluación, ya que los objetivos de pensamiento crítico implícitos o poco claros tienen un impacto mínimo o nulo en el aprendizaje. Los nuevos instrumentos permitirán a los docentes diferenciar mejor la instrucción, actualizar e innovar sus prácticas mediante estrategias que fomenten el pensamiento crítico (Kettler, 2014). Esto requiere que los profesores abandonen enfoques tradicionalistas, adopten métodos activos y reflexivos de enseñanza (Palma y Bores, 2002) e integren una formación más sólida en la enseñanza del pensamiento crítico durante su preparación profesional (Smetanová et al., 2015).

En síntesis, los antecedentes revisados muestran que, aunque el pensamiento crítico se reconoce como objetivo central de la formación escolar, su evaluación en educación primaria sigue siendo limitada, particularmente en el área matemática. La mayoría de los instrumentos disponibles fueron diseñados para niveles superiores, se apoyan en definiciones amplias o ambiguas del constructo, y rara vez consideran las características cognitivas de estudiantes de 7 a 10 años y el razonamiento inferencial informal. Esta combinación de alta demanda curricular y baja disponibilidad de herramientas válidas y pertinentes genera un vacío concreto en la medición del

pensamiento crítico en primer ciclo básico, que el presente estudio busca abordar mediante el diseño y validación de un instrumento específico para este contexto.

## 1.2 PLANTEAMIENTO DEL PROBLEMA

En abril de 2025, se realizó una revisión exhaustiva en la plataforma *Web of Science – Core Collection* utilizando el término de búsqueda "Measure critical thinking" en la opción "Topic". El criterio de selección incluyó artículos publicados a partir de 2014, tanto completos como de acceso anticipado. Para refinar los resultados, se aplicaron filtros en las categorías *Education Educational Research* y *Mathematics*, con el objetivo de identificar estudios centrados en la medición del pensamiento crítico en contextos educativos y en el área matemática. De un total de 1.388 artículos recuperados, se examinaron 1.000 trabajos directamente vinculados al tema. El análisis reveló que solo el 17% de estos abordaba la evaluación del pensamiento crítico en educación primaria, mientras que el 83% restante se concentraba en la educación superior. Esta distribución evidencia un vacío significativo en la investigación dirigida a las etapas escolares iniciales, lo cual resulta aún más notorio en lo que respecta al desarrollo y validación de instrumentos de evaluación adecuados para los primeros niveles educativos.

En la misma línea, Evendi et al., (2023), tras una revisión sistemática en la base de datos SCOPUS, reportaron 147 documentos sobre medición del pensamiento crítico. De ellos, solo el 6,4% correspondía al área de matemáticas, mientras que el 93,6% se centraba en otras disciplinas, principalmente ciencias sociales (35,2%), ingeniería (17,4%) y física y astronomía (11,9%). Este hallazgo confirma una clara carencia de instrumentos específicos para el contexto matemático. Esta problemática se agrava porque, como advierte Lai (2011), muchos de los instrumentos existentes no han pasado por procesos rigurosos de validación, lo que limita su confiabilidad y aplicabilidad en el aula.

Este vacío en la evaluación refleja, a su vez, una desconexión más profunda con la matemática inicial. Maričić y Špijunović (2015) lo ejemplifican al señalar que la mayoría de las conceptualizaciones del pensamiento crítico se han desarrollado en otros contextos,

con muy poca adaptación a los contenidos y las capacidades cognitivas de los estudiantes de primaria. Sin una base conceptual y valorativa sólida, resulta difícil superar los enfoques tradicionales que, según Gallegos et al. (2024), se basan en la memorización y la aplicación mecánica de procedimientos. Estos métodos reducen las oportunidades de desarrollar habilidades de orden superior, en contraste con la resolución de problemas contextualizados, que exige a los estudiantes analizar, evaluar y justificar sus soluciones.

Dado este panorama, se vuelve necesario ampliar la investigación en torno a la evaluación del pensamiento crítico en la educación primaria, con especial énfasis en muestras más representativas, distintos niveles escolares y diversos contextos. Es urgente, además, contar con instrumentos válidos y confiables que permitan evaluar esta competencia en estudiantes del primer ciclo escolar, considerando sus características cognitivas y el contexto disciplinar específico, como el área de matemáticas. Sin herramientas apropiadas, resulta difícil identificar las fortalezas y debilidades de los estudiantes en este ámbito, lo que limita la posibilidad de intervenciones pedagógicas efectivas y dificulta la toma de decisiones informadas por parte de docentes y equipos educativos.

Si la situación actual se mantiene, la evaluación del pensamiento crítico en matemáticas seguirá dependiendo de instrumentos no diseñados ni validados para estudiantes de 7 a 10 años, lo que impedirá identificar con precisión el nivel real de desarrollo de esta competencia. En consecuencia, los docentes continuarán tomando decisiones pedagógicas basadas en información parcial o poco fiable, se dificultará el diseño de intervenciones oportunas y efectivas, y las escuelas carecerán de evidencia sólida para monitorear el progreso de los estudiantes. Este escenario mantendrá las brechas diagnósticas actuales y limitará de manera significativa el impacto educativo del desarrollo del pensamiento crítico en la educación primaria.

### 1.2.1 Justificación del estudio

El pensamiento crítico es considerado un atributo esencial de los individuos del siglo XXI, debido a eso, los sistemas educativos han buscado integrarlo en sus currículos escolares, fortaleciendo las habilidades de pensamiento crítico de los estudiantes de los diferentes niveles de educación básica (Cañongo et al, 2020). La educación primaria, en particular, ofrece un momento clave para fortalecer estas habilidades, dado que los niños cuentan con la capacidad de razonar críticamente bajo condiciones educativas adecuadas (Lin et al., 2017).

A pesar de su relevancia, la mayoría de las investigaciones sobre pensamiento crítico se centran en la educación superior, dejando en segundo plano a los estudiantes más jóvenes, quienes presentan una mayor maleabilidad cognitiva para desarrollar estas habilidades (Deng et al., 2023). Esta brecha se manifiesta en la escasez de instrumentos diseñados específicamente para evaluar el pensamiento crítico en alumnos de educación primaria. Estos autores indican que, en varios contextos, no existen herramientas adaptadas a este grupo etario, lo que dificulta una evaluación precisa de sus habilidades.

Además, Choque (2019) advierte que existe una preocupación frecuente entre docentes respecto a la ausencia de instrumentos apropiados para niños, lo que ha llevado al uso inadecuado de pruebas no diseñadas para este nivel, limitando así la identificación y promoción efectiva del pensamiento crítico.

La falta de atención al desarrollo del pensamiento crítico en el ámbito matemático se ve agravada por la presencia de altos niveles de ansiedad matemática (Ramírez et al., 2018), que llevan a los estudiantes a evitar estrategias de recuperación de memoria y a depender de procedimientos mecánicos y menos eficientes (Imbo y Vandierendonck, 2007), y por la predominancia de metodologías centradas en la repetición mecánica. Estas prácticas restringen el desarrollo de habilidades superiores como la inferencia, la evaluación y la autorregulación (Güner y Gökçe, 2021).

A lo anterior se suma la dificultad de definir con precisión los componentes del pensamiento crítico, dada la diversidad de enfoques existentes y la falta de consenso respecto a su estructura conceptual (Ortega-Quevedo et al., 2020). Esta ambigüedad complica aún más la construcción de instrumentos validados, confiables y pertinentes para su evaluación en contextos escolares.

Frente a este panorama, la presente investigación busca explorar fundamentos teóricos y metodológicos que orienten la selección, adaptación o diseño de un instrumento de evaluación del pensamiento crítico adecuado para el primer ciclo de enseñanza básica, específicamente para estudiantes entre 7 a 10 años en el área de Matemáticas.

### **1.3 PREGUNTAS Y OBJETIVOS DE INVESTIGACIÓN**

#### 1.3.1 Preguntas de investigación

- A. ¿Qué aspectos conceptuales componen el pensamiento crítico?
- B. ¿Cuáles son los instrumentos o metodologías empleadas para evaluar el pensamiento crítico en matemáticas en estudiantes de primer ciclo, y qué características y limitaciones se identifican en ellos?
- C. ¿Qué criterios deben considerarse para diseñar un instrumento que evalúe de manera válida y pertinente el pensamiento crítico en matemáticas en el contexto escolar chileno?
- D. ¿Cómo contribuyen el juicio de expertos y las entrevistas cognitivas con estudiantes a establecer la validez de contenido del instrumento diseñado para evaluar el pensamiento crítico matemático en estudiantes de 7 a 10 años?

Estas preguntas de investigación son de carácter descriptivo e instrumental, ya que se orientan a identificar el estado actual de la evaluación del pensamiento crítico en matemáticas en educación primaria y, al mismo tiempo, a guiar el proceso de diseño y validación de un instrumento específico.

### 1.3.2 Objetivos de investigación

#### **Objetivo general:**

Desarrollar y realizar la validación de contenido de un instrumento de evaluación cuantitativo orientado a medir el pensamiento crítico en el área de matemáticas en estudiantes de entre 7 y 10 años.

#### **Objetivos específicos:**

OE1. Identificar los instrumentos existentes para la medición del pensamiento crítico en el contexto educativo, particularmente en la educación primaria.

OE2. Construir un instrumento cuantitativo para evaluar el pensamiento crítico matemático en estudiantes chilenos de 7 a 10 años.

OE3. Evaluar la validez del contenido del instrumento diseñado mediante el juicio de expertos y el análisis cualitativo de entrevistas realizadas a estudiantes.

Para responder a los objetivos planteados, se desarrolló un instrumento inicial basado en el modelo de habilidades y sub-habilidades del pensamiento crítico propuesto por Facione (1990; 2020). Este instrumento fue sometido a entrevistas cognitivas con estudiantes y a evaluaciones con jueces expertos mediante rúbricas con criterios previamente establecidos, medidos en escala Likert, lo que permitió obtener valores cuantitativos y observaciones cualitativas. Con los puntajes entregados, se realizaron ajustes a los ítems del instrumento, aplicando procedimientos de análisis. Este proceso permitió avanzar hacia el cumplimiento de los objetivos.

### 1.3.3 Naturaleza del estudio

Este estudio presenta un enfoque mixto, con diseño no experimental transeccional descriptivo, ya que no se manipulan variables y los datos se obtienen en un único momento temporal. Este tipo de diseño permite recopilar la información sobre cada uno de los criterios establecidos para el análisis del instrumento y describir las características de la

validez de contenido en relación con la evaluación de las habilidades y sub-habilidades de pensamiento crítico en niños de 7 a 10 años.

Finalmente, este trabajo se organiza en seis capítulos. El primer capítulo corresponde a la introducción, en donde se presentan los antecedentes del estudio, planteamiento del problema, preguntas y objetivos y la naturaleza del diseño. El segundo capítulo desarrolla el marco conceptual, en donde se presentan los conceptos fundamentales relacionados con el pensamiento crítico. El tercer capítulo describe el marco metodológico la cual va detallando el diseño de investigación, los participantes, los instrumentos y procedimientos de análisis. El cuarto capítulo presenta el análisis y los resultados obtenidos a partir de las entrevistas cognitivas y las valoraciones de los jueces expertos. El quinto capítulo corresponde a la discusión de los hallazgos en relación con nuestros resultados y la literatura revisada. Por último, el sexto capítulo describe las conclusiones del estudio, limitaciones y proyecciones para futuras investigaciones. El estudio finaliza con las referencias bibliográficas y los anexos.

## II. MARCO CONCEPTUAL

### 2.1 DEFINICIÓN DEL PENSAMIENTO CRÍTICO

Inicialmente, definir de manera precisa el pensamiento crítico no resultaba sencillo, ya que distintos autores lo han abordado desde perspectivas diversas, coincidiendo en algunos aspectos y diferenciándose en otros. John Dewey (1910) señala que la acción del pensamiento es un proceso activo y organizado, clave para actuar de forma inteligente y no dejarse llevar por impulsos sin reflexión. Para Ennis (1962) el pensamiento crítico es “la evaluación correcta de las afirmaciones” (pág. 83). Según este autor, pensar críticamente implica juzgar con precisión el significado, la claridad, la coherencia, la relevancia, la confiabilidad y la justificación de diferentes tipos de declaraciones o afirmaciones. Por su parte, Lipman (1998) plantea que el pensamiento crítico es un proceso que se corrige a sí mismo, fundamentado en criterios claros y conscientes del contexto. Este proceso permite reconocer y afrontar prejuicios, se complementa con el pensamiento creativo, integra perspectivas individuales y colectivas, y resulta fundamental en la educación para formar personas con juicio y capacidad de razonamiento.

En las últimas décadas, el pensamiento crítico ha sido entendido como un concepto flexible, aplicable a diferentes contextos y disciplinas, con especial relevancia en el ámbito educativo (Uribe-Hincapié y Gutiérrez-Ríos, 2023). Ennis (2018) lo redefine como “el pensamiento reflexivo y razonable centrado en decidir qué creer o hacer”, destacando un núcleo común en todas las definiciones: la toma de decisiones fundamentadas y la evaluación crítica de argumentos.

Desde una perspectiva más actual y operativa, Deng et al. (2023) y Facione (2020) señalan que el pensamiento crítico combina habilidades cognitivas y disposiciones personales —como la búsqueda de la verdad, la apertura mental y la sistematicidad—, permitiendo analizar objetivamente situaciones y formular juicios bien sustentados.

Willingham (2008) añade que implica considerar distintos puntos de vista, aceptar evidencia contradictoria, razonar con imparcialidad, exigir fundamentos empíricos e inferir conclusiones lógicas, alejándose del pensamiento irracional o meramente memorístico.

En el campo de las matemáticas, se ha definido como la capacidad de aplicar un razonamiento lógico para aprender conceptos, tomar decisiones y resolver problemas (Güner y Gökçe, 2021), lo que incluye comparar información, seleccionar datos relevantes y evaluarlos de forma crítica (Sosa-Gutiérrez et al., 2024).

Para esta investigación se adopta la definición de pensamiento crítico propuesta por Facione (1990), formulada mediante un consenso de expertos: “un juicio autorregulado y con propósito, que da como resultado la interpretación, el análisis, la evaluación y la inferencia, así como la explicación de las consideraciones evidenciables, conceptuales, metodológicas, criteriológicas o contextuales en las que se basa ese juicio” (p. 3). Esta concepción integra tanto las dimensiones cognitivas como las disposiciones necesarias para un pensamiento reflexivo, sistemático y fundamentado, resultando especialmente pertinente para su estudio en el ámbito del aprendizaje de las matemáticas (Abrami et al., 2008).

## **2.2 COMPONENTES DEL PENSAMIENTO CRÍTICO**

Diversos autores han coincidido en que el pensamiento crítico no se limita a un conjunto de habilidades cognitivas, sino que también incluye disposiciones personales que orientan y motivan su ejercicio (Facione, 2020; Ennis, 2018). Por ello, se propone comprender el pensamiento crítico desde una estructura tridimensional compuesta por habilidades, sub-habilidades y disposición.

### **2.2.1 Habilidades del pensamiento crítico**

Para que los estudiantes desarrollen estas habilidades, es fundamental que su desarrollo reciba importancia mediante una instrucción intencional y enfocada, ya que los

enfoques de enseñanza actuales han demostrado ser insuficientes. En el caso de las matemáticas, disciplina sustentada en el razonamiento lógico, la formulación de suposiciones, la generalización y la toma de decisiones, el pensamiento crítico y la disposición hacia él influyen tanto en el plano cognitivo como en el afectivo de los procesos de aprendizaje (Güner y Gökçe, 2021).

Según el Informe Delphi del Proyecto Americano sobre Pensamiento Crítico, y su reformulación posterior (Facione, 1990; 2020), las habilidades esenciales de esta competencia son: interpretación, análisis, evaluación, inferencia, explicación y autorregulación.

- **Interpretación:** De acuerdo con Casas et al. (2005), consiste en emitir juicios y argumentaciones fundamentadas que integren conocimientos científicos con convicciones personales, ideológicas y sociales. Esta habilidad permite analizar los hechos desde diversas perspectivas, evitar posturas dogmáticas y fomentar tanto el pensamiento crítico como el compromiso social. Lin et al. (2017) la asocian con la comprensión de la información recibida y la capacidad de formular y expresar ideas relacionadas. Facione (Informe Delphi, 1990; 2020) la define como la capacidad de captar y comunicar el sentido de experiencias, situaciones, datos, juicios, creencias, normas o procedimientos.
- **Análisis:** Según Elder y Paul (2019), es la capacidad de descomponer de manera sistemática cualquier pensamiento o razonamiento en sus elementos esenciales, para comprender su lógica interna y evaluar críticamente los procesos de razonamiento. Implica identificar relaciones inferenciales, tanto previstas como reales (Lin et al., 2017), y reconocer conexiones lógicas entre enunciados, conceptos o descripciones que reflejan creencias, juicios, experiencias o información (Facione, 1990, Informe Delphi; Facione, 2020).
- **Evaluación:** Para Elder y Paul (2019), es el proceso de valorar la calidad del pensamiento propio y ajeno aplicando criterios intelectuales objetivos. Supone

juzgar la credibilidad de las afirmaciones y representaciones (Lin et al., 2017) y evaluar la solidez lógica de las relaciones inferenciales (Facione, 1990; 2020).

- **Inferencia:** McNamara (2004) la describe como un proceso cognitivo activo y estratégico que conecta información explícita e implícita con el conocimiento previo para lograr una comprensión más profunda. Implica identificar información relevante para extraer conclusiones y prever consecuencias (Lin et al., 2017). Facione (1990; 2020) añade que incluye reconocer y emplear los elementos necesarios para formular hipótesis, considerar información pertinente y deducir consecuencias a partir de datos, principios o juicios.
- **Explicación:** Según Jorba et al. (2000), consiste en presentar argumentos que muestren relaciones causales para comprender mejor los hechos y ampliar o modificar el conocimiento. Lin et al. (2017) la asocian con la justificación del razonamiento que sustenta los resultados, mientras que Facione (1990; 2020) resalta la importancia de exponer conclusiones de forma clara, coherente y fundamentada.
- **Autorregulación:** Whitebread y Basilio (2012) la definen como la capacidad de adaptarse a situaciones emocionalmente desafiantes, controlar conductas inapropiadas y favorecer aquellas socialmente adecuadas. Es una habilidad clave que permite optimizar el pensamiento crítico mediante el control consciente de las actividades cognitivas, evaluando los juicios inferenciales para confirmar, corregir o validar el razonamiento (Facione, 1990; 2020; Tamayo et al., 2015).

De manera complementaria, Ennis (2018) propone una clasificación más amplia de las habilidades de pensamiento crítico, que abarca desde destrezas básicas, como la clarificación de ideas y la formulación de preguntas, hasta habilidades más complejas, como la evaluación de supuestos, el razonamiento deductivo e inductivo y la autorreflexión sobre la calidad del propio pensamiento.

## 2.2.2 Sub-habilidades del pensamiento crítico

Facione (2020) no sólo identifica un conjunto de habilidades centrales del pensamiento crítico, sino que además descompone cada una de ellas en sub-habilidades

**Tabla 2**

*Habilidades y sub-habilidades del pensamiento crítico de acuerdo con el modelo de Facione (1990; 2020)*

<b>A. Interpretación</b>	
A1. Categorización	Consiste en organizar la información en grupos o categorías para poder entenderla mejor. Sirve para describir experiencias, situaciones o ideas de una manera clara y con sentido.
A2. Decodificación del significado	Implica reconocer y comprender lo que se transmite en distintos sistemas de comunicación (lenguaje, gestos, gráficos, símbolos, etc.), captando tanto la información literal como las intenciones, valores, reglas o puntos de vista que están detrás del mensaje.
A3. Clarificación del sentido	Se refiere a explicar con claridad el significado de palabras, ideas o representaciones (como gráficos, símbolos o comportamientos), usando recursos como la descripción, la comparación o ejemplos. Su propósito es eliminar la confusión, la ambigüedad o la falta de precisión en lo que se quiere comunicar.
<b>B. Análisis</b>	
B1. Examen de ideas	Supone analizar las expresiones o planteamientos dentro de un contexto de razonamiento o argumentación, para comprender su función. Incluye definir términos, comparar o contrastar ideas, identificar problemas y sus partes, así como reconocer las relaciones entre esos elementos y el conjunto.
B2. Detección de argumentos	Consiste en identificar si un conjunto de afirmaciones, preguntas o representaciones busca dar razones a favor o en contra de una idea, opinión o punto de vista.
B3. Análisis de argumentos	Implica descomponer un razonamiento para reconocer sus partes: la conclusión principal, las premisas que la apoyan, los supuestos no expresados, las conclusiones intermedias y la estructura general del argumento, diferenciándolos de los elementos que no forman parte del razonamiento.
<b>C. Evaluación</b>	
C1. Valoración de la credibilidad	Se refiere a evaluar si una fuente de información u opinión es confiable, considerando factores relevantes y el contexto. También implica juzgar qué tan aceptables o verídicas son las experiencias, juicios o creencias que se presentan.

C2. Valoración de los argumentos	Implica evaluar un argumento y decidir si sus razones realmente sostienen la conclusión. También incluye fijarse en supuestos débiles o falsos, pensar en posibles objeciones y ver si nueva información haría que el argumento sea más fuerte o más débil.
<b>D. Inferencia</b>	
D1. Consulta de pruebas	Consiste en identificar qué fundamentos necesitan respaldo y diseñar estrategias para buscar información que lo proporcione. También implica valorar la relevancia de la información para decidir sobre la aceptabilidad de una afirmación, teoría o problema, y determinar formas adecuadas de investigación para obtenerla.
D2. Conjetura de alternativas	Implica proponer diferentes opciones para responder a una pregunta o resolver un problema. Supone elaborar hipótesis o planes alternativos, identificar supuestos y anticipar las posibles consecuencias de decisiones, teorías, políticas o creencias.
D3. Extracción de conclusiones	Consiste en usar la lógica y distintos tipos de razonamiento para decidir qué conclusión está mejor apoyada por la información y descartar las menos sólidas.
<b>E. Explicación</b>	
E1. Descripción de resultados	Implica expresar con precisión las conclusiones obtenidas del razonamiento, para poder analizarlas, evaluarlas o revisarlas.
E2. Justificación de procedimientos	Se centra en explicar los pasos seguidos en un razonamiento, mostrando la evidencia, conceptos, métodos o criterios usados, para que el proceso pueda ser comprendido, evaluado o mejorado.
E3. Presentación de argumentos	Consiste en dar razones que respalden una afirmación y explicar cómo se manejan posibles objeciones sobre el razonamiento o la evidencia utilizada.
<b>F. Autorregulación</b>	
F1. Autoexamen	Se enfoca en evaluar críticamente el propio razonamiento, identificando posibles sesgos o errores, y verificando que las conclusiones y procesos sean objetivos, racionales y justificados.
F2. Autocorrección	Busca diseñar estrategias y procedimientos para corregir errores o deficiencias identificadas en el propio razonamiento y sus causas.

*Nota.* La presente tabla se basa en el modelo de pensamiento crítico propuesto por Facione (1990; 2020). Las habilidades y sub-habilidades descritas han sido sintetizadas para efectos de claridad metodológica.

### 2.2.3 Disposiciones al pensamiento crítico

Un aspecto fundamental del pensamiento crítico es la disposición para ponerlo en práctica, definida como la motivación intrínseca para enfrentar problemas y tomar

decisiones basadas en la reflexión. Esta disposición se manifiesta en actitudes como ver ambos lados de un tema (Willingham, 2008), estar abierto a nueva evidencia que contradiga las ideas propias, y exigir que las afirmaciones estén respaldadas por evidencia (Facione, 1990).

Sin embargo, cuando se separan las habilidades y las disposiciones en la concepción del pensamiento crítico, se percibe cierta ambigüedad respecto al concepto mismo de disposición. Esto ocurre porque a veces se entiende exclusivamente como motivación, o bien se le asigna un papel menor a lo actitudinal dentro de la formación escolar, lo que limita su potencial formativo (Uribe-Hincapié y Gutiérrez-Ríos, 2023).

Facione (2020) describe al pensador crítico ideal no sólo en términos de habilidades cognitivas, sino también por su forma de enfrentar la vida.

En su propuesta original, Facione (1990) presentó una clasificación detallada de las disposiciones afectivas del pensamiento crítico, agrupadas en dos enfoques: hacia la vida en general y hacia problemas específicos.

Asimismo, se destaca que una persona comprometida con el pensamiento crítico busca la verdad con objetividad, integridad y ecuanimidad, evitando conductas explotadoras o prejuiciosas que podrían debilitar este proceso. Por lo tanto, el pensamiento crítico no implica sólo destrezas y sub-destrezas cognitivas, sino también una motivación interna que impulsa a actuar de manera más reflexiva y ética frente a los problemas (Facione, 2020).

Complementariamente, Ennis (2018) sostiene que los pensadores críticos ideales se caracterizan por una serie de disposiciones que refuerzan el uso responsable y riguroso de sus habilidades cognitivas. Estas incluyen la búsqueda de declaraciones claras sobre las conclusiones o preguntas, el esfuerzo por ofrecer razones explícitas y comprender cómo se relacionan entre sí, el deseo de estar bien informados y el uso de fuentes y observaciones creíbles, que habitualmente deben ser mencionadas. Además, valora la

consideración del contexto completo y la atención al problema principal en una situación determinada.

También enfatiza la importancia de estar alerta frente a posibles alternativas, mantener la mente abierta, considerar seriamente otros puntos de vista y abstenerse de emitir juicios cuando no se dispone de pruebas o razones suficientes. El pensador crítico ideal, según Ennis (2018), debe ser capaz de adoptar o cambiar de postura cuando las razones y evidencias lo justifican, buscar precisión en la medida que el tema lo permita, y, en general, esforzarse por alcanzar la verdad cuando tiene sentido hacerlo.

Tamayo et al. (2015) describen a los pensadores críticos como personas que: a) intentan identificar los supuestos que subyacen a las ideas, creencias, valores y acciones, b) poseen la capacidad de imaginar y explorar alternativas a formas existentes de pensar y vivir; c) son escépticos ante afirmaciones de verdades universales o explicaciones definitivas; y d) están conscientes del contexto en el que se producen y evalúan las ideas.

### **2.3 CONTEXTUALIZACIÓN DE LA INFERENCIA ESTADÍSTICA**

En el ámbito de la Educación Estadística, el estudio de la inferencia es reciente. Se la considera la culminación de la Estadística por su utilidad para fundamentar decisiones, juicios y análisis en investigaciones (Lugo et al., 2020).

En sentido amplio, la inferencia estadística es un tipo de razonamiento inferencial, definido como el proceso y el resultado de crear o probar generalizaciones que son de naturaleza probabilística a partir de los datos. Esta inferencia se constituye en el medio para formular conclusiones bajo incertidumbre cuando únicamente se dispone de datos parciales (Rubin et al., 2006).

La inferencia estadística, en su concepto más amplio, implica tomar los datos que se tienen a mano y, a partir de ellos, deducir algo acerca de un conjunto mayor del que no se tiene información completa (Gioria, 2015). Pfannkuch (2006) define la inferencia estadística informal como la habilidad de sacar conclusiones a partir de datos, basándose

principalmente en observar y comparar. Por su parte, la inferencia se refiere al conjunto de teorías y procedimientos que permiten tomar decisiones o hacer afirmaciones sobre una población, casi siempre basándose en datos obtenidos mediante un muestreo aleatorio (Gioria, 2015).

### 2.3.1 La necesidad del enfoque informal

Debido a que la inferencia estadística sintetiza conceptos clave como gráficos, promedios, variabilidad y muestreo, abordarla de modo informal y de forma recurrente permite a los estudiantes construir gradualmente la base cognitiva indispensable para el razonamiento inferencial (Zieffler et al., 2008).

Puesto que los estudiantes jóvenes no pueden acceder aún a los conceptos y procedimientos de la inferencia formal, es imprescindible recurrir a un enfoque didáctico de carácter informal (Ben-Zvi, 2006).

La inferencia estadística informal funciona como un puente entre el simple análisis exploratorio de datos y la inferencia estadística formal. Los estudios indican que, si se introduce este tipo de razonamiento desde las primeras etapas educativas, los estudiantes podrán abordar los métodos formales con mayor facilidad y menos dificultad (Gioria, 2015). Los nuevos currículos de estadística apuestan por priorizar la inferencia informal sobre los procedimientos formales, introduciéndola pronto, repasándola con frecuencia y aprovechando herramientas tecnológicas y simulaciones para su desarrollo (Ben-Zvi, 2006).

### 2.3.2 Definición del razonamiento inferencial informal (RII)

El Razonamiento Inferencial Informal (RII) es un concepto relativamente reciente en la literatura de investigación y ha recibido varias definiciones (Rodríguez, 2013). Se entiende por conocimiento estadístico informal al conjunto de ideas, intuiciones y experiencias previas que los estudiantes poseen acerca de los datos, las muestras o la incertidumbre antes de estudiar la inferencia estadística formal. Este conocimiento, que

emerge de situaciones de la vida cotidiana, constituye el punto de partida para el desarrollo de un razonamiento inferencial intuitivo (Lugo et al., 2020).

Otros autores han contribuido a definir el RII:

(1) Rubin, Hammerman y Konold (2006) definieron el razonamiento inferencial informal como un enfoque preliminar que se basa en comprender propiedades de los conjuntos de datos (como distinguir patrones sistemáticos (señal) de la variabilidad natural (ruido)), reconocer el papel del tamaño de la muestra y del muestreo aleatorio para evitar sesgos, y diferenciar entre afirmaciones que son siempre verdaderas y aquellas que solo lo son a menudo, todo ello sin recurrir aún a cálculos probabilísticos formales.

(2) Pfannkuch (2006) caracteriza la inferencia informal como la habilidad de sacar deducciones apoyándose fundamentalmente en la observación y el análisis comparativo de los patrones que muestran los gráficos estadísticos.

(3) Zieffler, Garfield, DelMas y Reading (2008) proporcionaron una definición de RII que es ampliamente utilizada, estableciéndose como la habilidad de los estudiantes para construir argumentos —apoyándose en su conocimiento estadístico intuitivo— que les permitan extraer conclusiones acerca de una población mayor a partir de la evidencia proporcionada por una muestra observada, sin recurrir aún a métodos estadísticos formales.

### 2.3.3 El RII como un proceso de razonamiento

Según Zieffler, Garfield, DelMas y Reading (2008), un marco para el estudio del razonamiento inferencial informal debe articularse en torno a tres componentes clave:

1. La capacidad de formular afirmaciones o conjeturas sobre poblaciones a partir de información muestral, evitando recurrir a técnicas estadísticas formales.
2. La habilidad para movilizar y sintetizar tanto saberes previos formales como conocimientos informales disponibles.

3. La competencia para fundamentar dichas afirmaciones mediante argumentos explícitos basados en la evidencia de los datos.

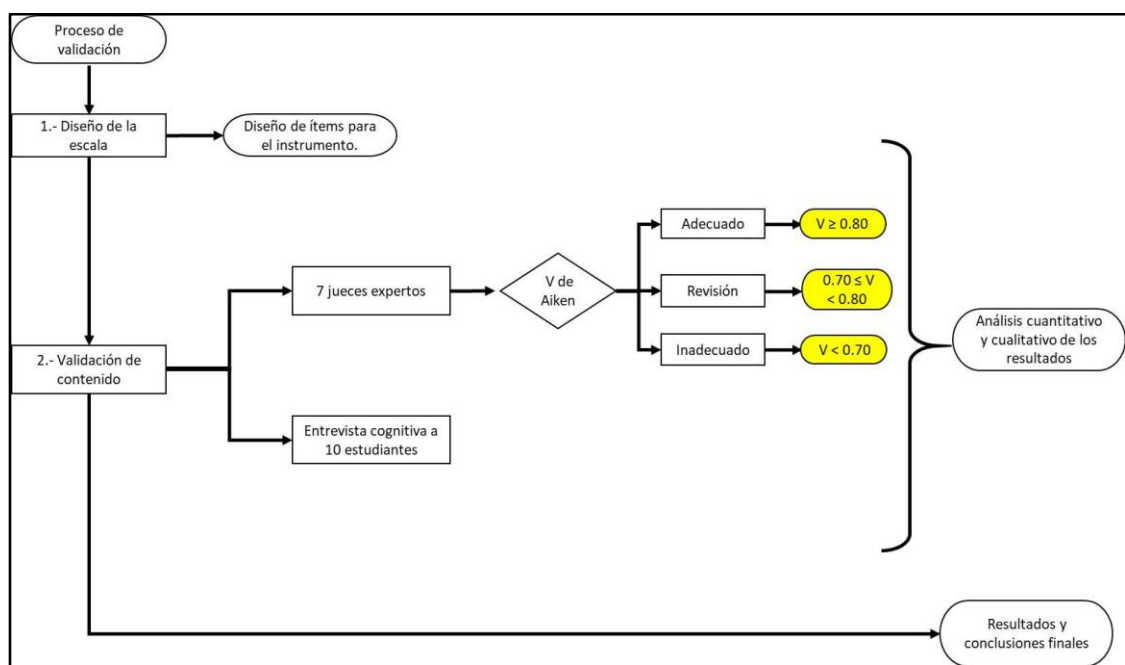
De manera más detallada, el RII puede resumirse como un proceso que incluye (Rodríguez, 2013): (1) inferir características de una población a partir de lo observado en una muestra, como su tendencia central o distribución, (2) evaluar si las diferencias entre dos grupos son significativas o simplemente producto del azar, comparando sus muestras, y (3) juzgar qué tan probable o inesperado es un resultado muestral, dadas ciertas expectativas previas.

En síntesis, el marco conceptual establece una comprensión detallada y estructurada del pensamiento crítico, integrando sus definiciones, habilidades, sub-habilidades y disposiciones desde enfoques clásicos y actuales. Este cuerpo teórico se complementa y enriquece con la incorporación del Razonamiento Inferencial Informal (RII) como componente cognitivo clave, entendido como la habilidad para construir argumentos y extraer conclusiones sobre una población a partir de evidencia muestral, movilizand o saberes previos e intuitivos sin recurrir aún a métodos estadísticos formales. La explicitación de estas dimensiones (tanto las actitudinales del pensamiento crítico como las procedimentales del RII) fundamenta la pertinencia de construir un instrumento específico para educación primaria y proporciona los lineamientos conceptuales necesarios para su diseño, operacionalización y validación. Con ello, el capítulo sienta las bases teóricas que articulan el problema de investigación con las decisiones metodológicas desarrolladas en los capítulos siguientes.

### III. METODOLOGÍA

La presente investigación busca garantizar un proceso metodológico coherente con el objetivo general. Para ello, se propone una metodología estructurada en diversas etapas, como se muestra en la Figura 1.

**Figura 1**  
*Esquema sobre la metodología de la investigación*



*Nota.* Elaboración propia.

#### 3.1 ENFOQUE Y DISEÑO DE LA INVESTIGACIÓN

Este estudio adopta un enfoque de métodos mixtos de tipo integración múltiple, que implica combinar los métodos cuantitativo y cualitativo en un mismo estudio para abordar el problema de investigación de manera más completa (Hernández-Sampieri y Mendoza-Torrez, 2023), orientado al diseño y validación de un instrumento que permita evaluar el pensamiento crítico en estudiantes de 7 a 10 años, específicamente en el área de matemática. Para esto, se utiliza un diseño iterativo, que funciona por ciclos repetidos

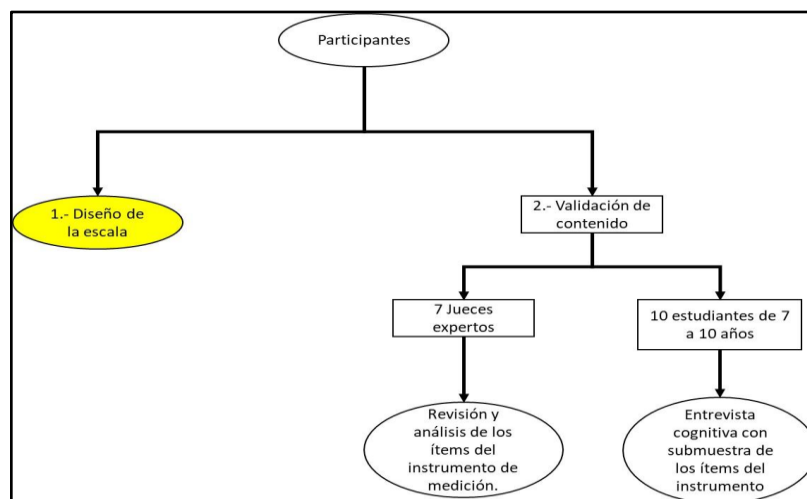
de planificación, recolección de datos, análisis y ajuste, en lugar de seguir un plan rígido y lineal. Este proceso permite adaptar las preguntas, métodos o instrumentos sobre la marcha, profundizando progresivamente en el tema hasta alcanzar una comprensión sólida o hasta que nueva información deje de aparecer (Hernández-Sampieri et al., 2014).

Este tipo de diseño facilita la recolección de información sobre fenómenos en contextos naturales, entregando una descripción detallada de la realidad sin intervención directa por parte del investigador (Hernández-Sampieri et al., 2014). Además, responde a la necesidad de diseñar instrumentos específicos para medir habilidades cognitivas complejas en etapas tempranas, como el pensamiento crítico en matemáticas, donde actualmente existen limitaciones de herramientas adaptadas a dicha población. Esto se alinea con lo planteado por Vásquez (2022), quien resalta la importancia de desarrollar modelos e instrumentos adecuados que permitan evaluar y potenciar el desarrollo del pensamiento crítico desde la educación primaria.

### 3.2 PARTICIPANTES

Como se muestra en la Figura 2, los participantes de esta investigación se dividen en dos grupos principales: estudiantes y jueces expertos.

**Figura 2**  
*Esquema metodológico sobre los participantes*



*Nota.* Elaboración propia.

La selección de los 10 estudiantes se realizará mediante un muestreo no probabilístico de tipo intencional o dirigido. Esta estrategia se elige porque el estudio se encuentra en una fase inicial exploratoria y de refinamiento del instrumento, donde el interés primordial reside en la riqueza, profundidad y calidad de la información respecto a la comprensión y significado de los ítems, más que en la cantidad o en la generalización estadística (Hernández-Sampieri et al., 2014).

El tamaño de 10 casos es sugerido para los estudios intensivos o para los grupos de enfoque. Esta pequeña muestra se ajusta a un muestreo que no busca generalizar resultados a la población en esta etapa, sino obtener la información cuantitativa y cualitativa necesaria para la modificación y ajuste de los ítems antes de la fase de validación de contenido formal (Muñiz y Fonseca-Pedrero, 2019).

Se establecen los siguientes criterios de inclusión: deberán ser niños y niñas que cursen entre segundo y cuarto año básico, con edades entre 7 y 10 años en las regiones del Biobío y La Araucanía. Estos criterios cumplen con la necesidad de delimitar claramente la población según sus características de contenido, lugar y tiempo (Hernández-Sampieri et al., 2014; Muñiz y Fonseca-Pedrero, 2019).

Se priorizará la participación de estudiantes pertenecientes a establecimientos educacionales municipales y particulares subvencionados para reflejar la diversidad socioeconómica y respetar el principio de equidad educativa (OECD, 2019). Este enfoque en la diversidad es fundamental, dado que la segregación socioeconómica de la población escolar chilena es comparativamente alta a nivel internacional, lo que afecta diferencialmente las oportunidades y logros de aprendizaje de los alumnos (Bellei, 2013). Por ello, se procurará diversidad en cuanto a género, tipo de establecimiento y contextos socioeconómicos, con énfasis en sectores urbanos.

Finalmente, se establece como criterio de exclusión la presencia de Necesidades Educativas Especiales (NEE) de carácter permanente que puedan afectar

significativamente la comprensión o respuesta al instrumento. Este criterio busca que las variables del estudio sean observables y medibles, con el fin de minimizar el error de medición y asegurar la calidad de las respuestas en esta fase crítica de refinamiento del instrumento (Hernández-Sampieri et al., 2014; Muñiz y Fonseca-Pedrero, 2019).

Asimismo, se contará con la colaboración de siete jueces expertos, seleccionados mediante un muestreo no probabilístico intencional para conformar el panel de validación, ya que este tipo de muestreo permite al investigador seleccionar sujetos según criterios específicos que los hacen idóneos para el estudio, particularmente cuando se requiere la opinión de expertos para procesos de validación de instrumentos (Hernández-Sampieri et al., 2014). Los jueces deberán cumplir con los siguientes criterios: poseer formación y experiencia comprobada en psicología educativa y didáctica de la matemática, con un mínimo de cinco años de ejercicio profesional en el área; además, deberán mantener imparcialidad, siendo independientes del equipo de investigación para evitar posibles sesgos (Lynn, 1986; Muñiz y Fonseca-Pedrero, 2019). La decisión de incluir a siete expertos se fundamenta en lo planteado por Lynn (1986), quien recomienda la participación de entre cinco y diez jueces para asegurar una evaluación equilibrada, señalando que un número excesivo puede dificultar el consenso, mientras que una cantidad muy reducida compromete la confiabilidad del proceso.

La elección del segundo, tercero y cuarto año básico como niveles de aplicación responde tanto a consideraciones pedagógicas como evolutivas. Desde lo pedagógico, estos niveles marcan una etapa clave en la consolidación de habilidades numéricas, estrategias de resolución de problemas y razonamiento matemático, fundamentales para el desarrollo de un pensamiento crítico y autónomo en el ámbito matemático (MINEDUC, 2024). Desde lo cognitivo, los estudiantes en los primeros años de la educación básica ya evidencian capacidades emergentes de pensamiento crítico, especialmente cuando son expuestos a contextos de enseñanza enfocados en la reflexión y el cuestionamiento (O'Reilly et al., 2022). Por tanto, evaluar el pensamiento crítico desde segundo hasta

cuarto básico resulta pertinente y estratégico para fomentar su desarrollo desde una etapa temprana.

### 3.3 DISEÑO Y CARACTERÍSTICAS DEL INSTRUMENTO

El instrumento propuesto para esta investigación corresponde a una prueba de respuesta cerrada de opción múltiple con única respuesta correcta, también conocida como test de opción múltiple monovalente (Hernández-Sampieri et al., 2014). Esta decisión se basa en que las pruebas de opción múltiple, cuando están bien diseñadas, son una herramienta efectiva para evaluar el razonamiento complejo y la capacidad de resolver problemas. La evidencia indica que este tipo de preguntas, en especial aquellas que presentan situaciones o casos que exigen analizar y aplicar el conocimiento, pueden medir eficazmente las habilidades de pensamiento crítico (Zaidi et al., 2018).

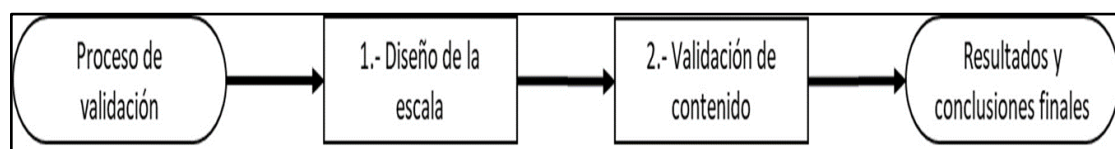
En coherencia con la definición conceptual presentada, el pensamiento crítico se operacionaliza en este instrumento como un constructo jerárquico compuesto por seis habilidades centrales (interpretación, análisis, evaluación, inferencia, explicación y autorregulación) y sus respectivas sub-habilidades. Cada sub-habilidad se representa mediante ítems de opción múltiple con una única respuesta correcta, cuyas respuestas se codifican de forma dicotómica (1 = correcta, 0 = incorrecta).

### 3.4 PROCEDIMIENTO DE VALIDACIÓN

El proceso de validación del instrumento diseñado para medir el pensamiento crítico en estudiantes de segundo a cuarto básico, en el contexto matemático, se organiza en cuatro etapas (Figura 3).

#### Figura 3

*Esquema sobre el proceso de validación del instrumento de medición*



*Nota.* Elaboración propia.

Cada una de estas etapas cumple una función específica y sigue un enfoque riguroso para asegurar la validez.

En la primera etapa, se elabora el instrumento de opción múltiple con una única respuesta correcta, a partir de una revisión teórica del pensamiento crítico. La formulación de los ítems se guía por las sub-habilidades descritas por Facione (2020) en el informe Delphi, utilizando el pensamiento crítico como un factor general, mientras que las habilidades específicas de pensamiento crítico constituirán los sub-factores. Se procurará que el lenguaje y el formato sean adecuados para estudiantes de 7 a 10 años, asegurando la coherencia con el currículo nacional de matemáticas.

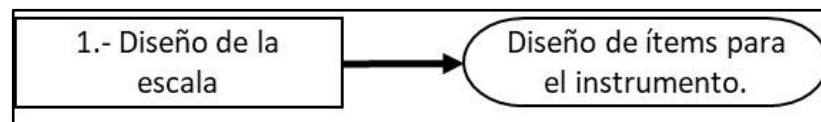
La segunda etapa corresponde a la validación de contenido mediante la evaluación de expertos y entrevistas cognitivas. Siete jueces expertos especialistas en psicología educativa y didáctica matemática evaluarán los ítems según seis criterios: claridad, pertinencia, contextualización para el razonamiento inferencial informal, relevancia de las alternativas y cobertura del pensamiento crítico por ítem, utilizando una escala tipo Likert. Además del análisis descriptivo de sus valoraciones, se aplicará un procedimiento mixto que integra el examen cualitativo de las observaciones y un proceso sistemático de identificación de ítems que requieran revisión profunda. Los ítems que presenten debilidades serán reformulados considerando las sugerencias de los jueces y posteriormente reevaluados para asegurar la pertinencia de los ajustes. En paralelo, se realizarán entrevistas cognitivas con diez estudiantes para verificar su comprensión de los ítems y complementar la información obtenida. Finalmente, las valoraciones definitivas se analizarán mediante el coeficiente V de Aiken, fortaleciendo la validez de contenido del instrumento.

### 3.4.1 Diseño de la escala

El diseño de la escala consideró la elaboración de los ítems del instrumento, como se observa en la Figura 4.

#### **Figura 4**

*Esquema del diseño de la escala del instrumento de medición*



*Nota.* Elaboración propia.

El instrumento propuesto consiste en una prueba de opción múltiple con única respuesta correcta, diseñada para evaluar estudiantes de 7 a 10 años. Este formato permite obtener resultados objetivos y estandarizados, lo que lo hace adecuado para contextos educativos diversos (OECD, 2019). Su diseño se fundamenta en el modelo de pensamiento crítico desarrollado mediante el consenso de expertos en el Informe Delphi (Facione, 1990) y sus actualizaciones posteriores (Facione, 2020).

El desarrollo de los ítems se basa en tres niveles conceptuales jerarquizados:

- **Pensamiento crítico:** Constructo general que se busca evaluar.
- **Habilidades centrales:** Dimensiones fundamentales del pensamiento crítico que agrupan distintos procesos cognitivos (interpretación, análisis, evaluación, inferencia, explicación y autorregulación) y orientan la organización del instrumento.
- **Sub-habilidades:** Componentes operativos de cada habilidad, que orientan la redacción de los ítems y permiten representar acciones observables del pensamiento crítico.

A nivel psicométrico, el instrumento adoptará una estructura jerárquica en la que:

- El **pensamiento crítico** actuará como factor general.

- Las **habilidades centrales** constituirán los sub-factores, cada una medida por los ítems asociados a sus sub-habilidades.
- Las sub-habilidades funcionarán como referencia teórica para la elaboración de ítems, pero no se modelarán como dimensiones independientes, ya que cada una se encuentra representada por un solo ítem dentro de su habilidad correspondiente.

Esta configuración permite mantener la coherencia teórica del modelo de Facione (Informe Delphi, 1990; 2020), a la vez que asegura la viabilidad psicométrica del análisis factorial confirmatorio en la población objetivo.

Se procurará que tanto el lenguaje como el formato de los ítems sean adecuados al nivel de desarrollo cognitivo de los estudiantes y se alineen con los objetivos del currículum nacional de matemáticas. Esta adecuación se justifica en las Bases Curriculares: Primero a Sexto Básico (MINEDUC, 2024), las cuales señalan que entre segundo y cuarto básico los estudiantes se encuentran en una etapa de transición desde el pensamiento concreto hacia formas más abstractas de razonamiento matemático. Por ello, el aprendizaje en estos niveles se apoya en el uso de representaciones gráficas, pictóricas e icónicas, que facilitan la comprensión de relaciones y patrones, junto con el uso de lenguaje cotidiano y contextos familiares. En coherencia con ello, los ítems del instrumento presentan situaciones próximas a la realidad de los niños (como encuestas, elecciones o comparaciones cotidianas) acompañadas de apoyos visuales y formuladas con vocabulario comprensible, evitando expresiones abstractas o ambiguas que puedan interferir con la interpretación del problema. Este enfoque favorece una lectura accesible, promueve el razonamiento inferencial informal y permite evaluar el pensamiento crítico en un marco coherente con el desarrollo cognitivo esperado para este rango etario.

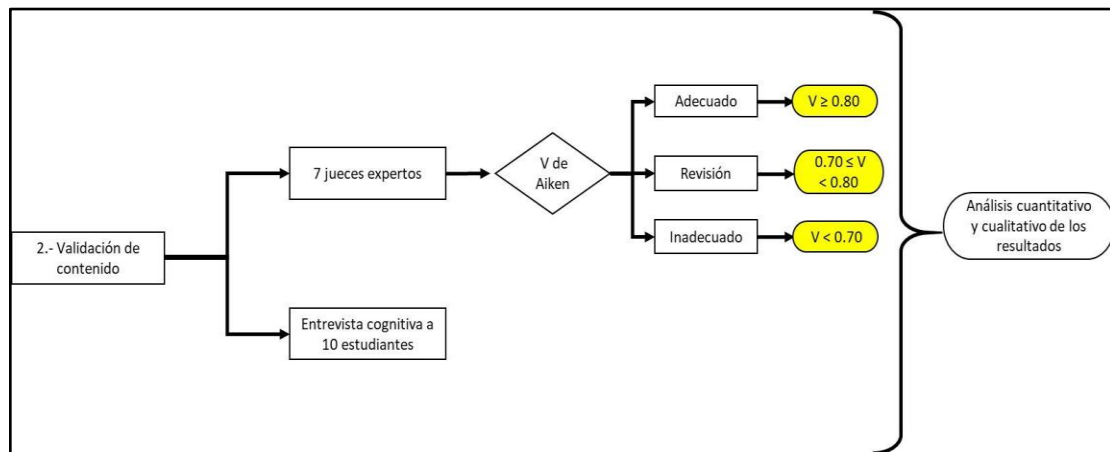
#### 3.4.2 Validación de contenido

Para garantizar la validez de contenido, se aplicará un proceso compuesto por dos fases complementarias. Por un lado, siete jueces expertos evaluarán cada ítem y, por otro lado, se realizarán entrevistas cognitivas a un grupo de diez estudiantes con sub-muestras

diferentes del instrumento (Figura 5). Ambas fases permitirán examinar el contenido desde perspectivas técnicas y desde la comprensión real de la población objetivo.

**Figura 5**

*Esquema sobre el proceso de validación de contenido del instrumento de medición*



*Nota.* Elaboración propia.

En coherencia con este enfoque, el proceso de validación incorporará una triangulación metodológica destinada a fortalecer el rigor del análisis. Esta triangulación considerará: (a) la combinación de métodos, integrando valoraciones cuantitativas mediante escalas Likert con evidencias cualitativas obtenidas en entrevistas cognitivas; (b) la diversificación de fuentes, al incluir tanto la evaluación de jueces expertos como la participación de estudiantes; y (c) la articulación de datos cuantitativos y cualitativos. Esta estrategia múltiple incrementa la validez y confiabilidad del procedimiento (Hernández-Sampieri et al., 2014).

### **Evaluación por jueces expertos**

Cada uno de los jueces expertos, evaluarán de manera independiente cada ítem mediante un instrumento de validación (ver [Anexo\\_1](#)), el cual contiene una presentación general del estudio, los objetivos de la investigación y el modelo teórico que sustenta el instrumento. Asimismo, se incorpora una tabla de síntesis conceptual del pensamiento

crítico, que describe las seis habilidades y sus sub-habilidades según Facione (1990, 2020), junto con su definición y cantidad de ítems asociados (Tabla 3).

**Tabla 3**  
*Síntesis conceptual del pensamiento crítico*

Pensamiento crítico			
El pensamiento crítico ha sido concebido como un proceso activo y reflexivo que guía la acción y la toma de decisiones (Dewey, 1910; Ennis, 2018), y como un proceso autocorrectivo, basado en criterios y consciente del contexto (Lipman, 1998). Para esta investigación, se adopta la definición de Facione (1990), que lo describe como un juicio autorregulado y con propósito, sustentado en habilidades cognitivas y disposiciones que permiten interpretar, analizar, evaluar e inferir de manera fundamentada (Facione, 2020).			
Habilidades	Sub-habilidades	Cantidad de ítems	Características generales
A. Interpretación	A1. Categorización	1	Comprender y comunicar el significado de información, experiencias o situaciones, integrando conocimientos científicos con perspectivas personales y sociales. Implica identificar, organizar y contextualizar datos o mensajes, reconociendo tanto sus significados explícitos como implícitos, para expresarlos de manera clara y coherente (Casas et al., 2005; Facione, 1990, Informe Delphi; Facione, 2020; Lin et al., 2017).
	A2. Decodificación del significado	1	
	A3. Clarificación del sentido	1	
B. Análisis	B1. Examen de ideas	1	Descomponer información, ideas o situaciones en sus elementos esenciales para comprender su lógica, evaluar razonamientos y reconocer relaciones inferenciales y conexiones entre conceptos. Facilita la identificación de premisas, conclusiones y suposiciones, así como la coherencia de los argumentos, promoviendo una comprensión
	B2. Detección de argumentos	1	

	B3. Análisis de argumentos	1	profunda y crítica de la información (Elder y Paul, 2019; Facione, 1990, Informe Delphi; Facione, 2020; Lin et al., 2017).
C. Evaluación	C1. Valoración de la credibilidad	1	Juzgar la calidad del pensamiento propio y ajeno, valorando la credibilidad de la información y la solidez lógica de los argumentos. Implica identificar premisas, conclusiones y supuestos, anticipar objeciones y determinar la aceptabilidad de afirmaciones, contribuyendo a razonamientos sólidos y decisiones fundamentadas (Elder y Paul, 2019; Facione, 1990, Informe Delphi; Facione, 2020; Lin et al., 2017).
	C2. Valoración de los argumentos	1	Juzgar la calidad del pensamiento propio y ajeno, valorando la credibilidad de la información y la solidez lógica de los argumentos. Implica identificar premisas, conclusiones y supuestos, anticipar objeciones y determinar la aceptabilidad de afirmaciones, contribuyendo a razonamientos sólidos y decisiones fundamentadas (Elder y Paul, 2019; Facione, 1990, Informe Delphi; Facione, 2020; Lin et al., 2017).
D. Inferencia	D1. Consulta de pruebas	1	Conectar información explícita e implícita con el conocimiento previo para extraer conclusiones fundamentadas. Implica formular hipótesis, generar alternativas y deducir consecuencias a partir de datos, principios o juicios, evaluando la plausibilidad de las opciones y fortaleciendo el razonamiento estratégico y la toma de decisiones basada en evidencia (Facione, 1990, Informe Delphi; Facione, 2020; McNamara, 2004; Lin et al., 2017).
	D2. Conjetura de alternativas	1	Conectar información explícita e implícita con el conocimiento previo para extraer conclusiones fundamentadas. Implica formular hipótesis, generar alternativas y deducir consecuencias a partir de datos, principios o juicios, evaluando la plausibilidad de las opciones y fortaleciendo el razonamiento estratégico y la toma de decisiones basada en evidencia (Facione, 1990, Informe Delphi; Facione, 2020; McNamara, 2004; Lin et al., 2017).
	D3. Extracción de conclusiones	1	Conectar información explícita e implícita con el conocimiento previo para extraer conclusiones fundamentadas. Implica formular hipótesis, generar alternativas y deducir consecuencias a partir de datos, principios o juicios, evaluando la plausibilidad de las opciones y fortaleciendo el razonamiento estratégico y la toma de decisiones basada en evidencia (Facione, 1990, Informe Delphi; Facione, 2020; McNamara, 2004; Lin et al., 2017).
E. Explicación	E1. Descripción de resultados	1	Presentar y comunicar de manera clara y fundamentada los resultados del razonamiento. Implica exponer relaciones causales, justificar procedimientos y ofrecer argumentos que respalden conclusiones, facilitando la comprensión de los hechos y la evaluación de su validez (Facione, 1990, Informe Delphi;
	E2. Justificación de procedimientos	1	Presentar y comunicar de manera clara y fundamentada los resultados del razonamiento. Implica exponer relaciones causales, justificar procedimientos y ofrecer argumentos que respalden conclusiones, facilitando la comprensión de los hechos y la evaluación de su validez (Facione, 1990, Informe Delphi;
	E3. Presentación de	1	Presentar y comunicar de manera clara y fundamentada los resultados del razonamiento. Implica exponer relaciones causales, justificar procedimientos y ofrecer argumentos que respalden conclusiones, facilitando la comprensión de los hechos y la evaluación de su validez (Facione, 1990, Informe Delphi;

argumentos		Facione, 2020; Jorba et al., 2000; Lin et al., 2017).	
F. Autorregulación	F1. Autoexamen	1	Supervisar y ajustar el propio razonamiento, reconocer errores o prejuicios, y corregirlos mediante estrategias apropiadas. Implica reflexionar sobre juicios, conclusiones y procedimientos, así como controlar conductas y emociones para mantener un pensamiento objetivo, coherente y eficaz (Facione, 1990, Informe Delphi; Facione, 2020; Tamayo et al., 2015; Whitebread y Basilio, 2012).
	F2. Autocorrección	1	

Posteriormente, como se muestra en la Tabla 4, el instrumento presenta una tabla de evaluación con escala tipo Likert de 4 puntos (1 = “Deficiente”, 2 = “Regular”, 3 = “Bueno”, 4 = “Excelente”) (Hernández-Sampieri et al., 2014), que contempla cinco criterios por ítem: claridad, pertinencia, contextualización para el razonamiento inferencial informal, relevancia de las alternativas y cobertura del pensamiento crítico por ítem. Cada puntaje de la escala está asociado a indicadores específicos que describen el grado de cumplimiento del criterio evaluado, permitiendo interpretar de manera precisa el nivel de adecuación de cada ítem y orientar las decisiones de mejora correspondientes.

**Tabla 4**  
*Ejemplo del criterio de “Claridad” presentado en “Instrumento de validación”*

Criterio	Pregunta orientadora	Puntaje	Indicador
Valoración por ítem			
Claridad	¿La redacción es clara y fácil de entender para niños de 7–10 años?	1	El ítem no se comprende. La estructura y el vocabulario son inadecuados, impidiendo entender la consigna. Necesita rediseño total

lenguaje claro y años, sin adecuado para ambigüedades? niños de 7-10 años.	2	El ítem se entiende con esfuerzo. Presenta ambigüedades o términos complejos que afectan su interpretación. Requiere reformulación.
	3	El ítem es claro en general, pero usa al menos un elemento (palabra o frase) que puede causar duda o que no es totalmente adecuado para la edad. Sugiere revisiones puntuales.
	4	El ítem es inmediatamente comprensible. El lenguaje y la estructura son precisos para la edad. No requiere modificaciones.

*Nota.* Ver tabla de evaluación completa en Anexo 1.

Cada juez experto recibirá una ficha de evaluación por ítem (Figura 6), la cual contiene información sobre el código del ítem, la habilidad y sub-habilidad que evalúa, junto con el enunciado correspondiente.

### Figura 6

*Ejemplo de ficha de evaluación por ítem*

ÍTEM N° #				
Código	Habilidad		Subhabilidad	
A1	Interpretación	Comprender y expresar el significado o importancia de una amplia variedad de experiencias, situaciones, datos, eventos, juicios, convenciones, creencias, reglas, procedimientos o criterios.	Categorización	Aprender a formular apropiadamente categorías, distinciones o marcos para comprender, describir o caracterizar la información. Describir experiencias, situaciones, creencias, acontecimientos, etc., de forma que adquieran significados comprensibles en términos de categorizaciones, distinciones o marcos apropiados.
[Pegar ítem aquí]				
Puntaje por criterio				
Claridad	Pertinencia	Contextualización para el razonamiento inferencial informal	Relevancia de las alternativas	Cobertura del pensamiento crítico por ítem
[Si considera necesario, incluya comentarios adicionales aquí]				

En esta ficha, el juez debe asignar puntajes a cada uno de los cinco criterios de valoración por ítem utilizando la escala Likert de cuatro puntos. Además, se incluye un espacio para observaciones cualitativas, donde los expertos pueden registrar sugerencias, comentarios o propuestas de mejora respecto al contenido o redacción del ítem.

### **Análisis de valoraciones obtenidas por la evaluación de jueces expertos**

Los datos recopilados serán analizados mediante un enfoque mixto, integrando el análisis cualitativo de las observaciones con un proceso sistemático de identificación y depuración de ítems (Hernández-Sampieri et al., 2014; Meliá, 2001).

#### **A. Identificación de ítems críticos.**

Para priorizar los ítems que requieren una revisión profunda, se aplicarán de manera concurrente los siguientes cuatro criterios de selección (Meliá, 2001). Un ítem será clasificado como crítico si cumple con al menos uno de ellos:

- **Criterio 1.** Acumulación de calificaciones bajas: Ítems que concentren entre 7 y 10 calificaciones bajas (puntajes 1 o 2 en la escala Likert) distribuidas entre los cinco criterios evaluados. Este patrón evidencia debilidades extendidas y multifacéticas.
- **Criterio 2.** Coincidencia entre jueces: Ítems donde tres o más jueces coincidan en asignarle una calificación baja en al menos un criterio. Esta convergencia señala un problema reconocido de manera consistente y no subjetiva.
- **Criterio 3.** Debilidad multifacética y concurrente: Ítems donde dos o más jueces le asignen calificaciones bajas en al menos dos criterios distintos. Esto refleja coincidencia entre evaluadores y amplitud en las dificultades detectadas.
- **Criterio 4.** Formación de grupos de debilidad: Ítems que presenten entre dos y tres "grupos de debilidad", definidos como la repetición de calificaciones bajas por parte de distintos jueces en un mismo criterio específico (ej.: tres jueces asignan puntuación baja en "claridad", formando un grupo).

Los ítems que no cumplan ninguno de estos criterios, o que presenten un máximo de 6 calificaciones bajas de forma aislada, serán considerados con un nivel de dificultad menor y podrán ser sometidos a ajustes menores o mantenerse (Meliá, 2001).

### **B. Proceso Cualitativo y Modificación de Ítems Críticos.**

Una vez identificados, los ítems críticos serán sometidos a un análisis cualitativo exhaustivo de las observaciones, sugerencias y comentarios específicos registrados por los jueces en sus fichas de evaluación. La reformulación de estos ítems incorporará de manera integral dichas sugerencias para abordar las debilidades sistémicas detectadas (Hernández-Sampieri et al., 2014; Meliá, 2014).

Para validar la efectividad de las modificaciones, las nuevas versiones de los ítems críticos se someterán a una reevaluación por parte de un subgrupo de los jueces originales, en particular aquellos cuyas observaciones motivaron los cambios. Esta evaluación seguirá utilizando la misma rúbrica Likert de 4 puntos.

### **C. Cálculo del coeficiente V de Aiken:**

Finalizado el ciclo de análisis y modificación, y con las valoraciones definitivas de todos los ítems (originales para los no modificados y nuevas para los reformulados), se procederá a la validación cuantitativa final con el coeficiente V de Aiken.

Este índice permitirá determinar el grado de acuerdo entre los expertos respecto a la relevancia de cada ítem (Aiken, 1980). La fórmula utilizada será:

$$V = \frac{\Sigma (ri - b)}{n \times (c - 1)}$$

**$\Sigma$** : Suma de los valores.

**ri**: Valor que emite cada juez ante un ítem.

**b**: Valor más bajo que se puede asignar en la escala de valoración.

**n:** Número de jueces expertos.

**c:** Valor máximo de valoración que se puede asignar en la escala de valoración.

En la Tabla 5, se establecen los criterios de decisión que orientarán la interpretación de los valores obtenidos para cada ítem.

**Tabla 5**

*Criterios de interpretación del coeficiente V de Aiken.*

Valores de V de Aiken	Interpretación
$0.80 \leq V$	Ítem adecuado. Se mantiene en el instrumento.
$0.70 \leq V < 0.80$	Ítem en revisión. Se reformula considerando observaciones de los jueces.
$V < 0.70$	Se elimina, salvo justificación teórica o empírica sólida.

*Nota.* Los criterios de interpretación fueron basados en Aiken (1980), Jiménez-Herrera (2025), Penfield y Giacobbi (2004).

Los resultados cuantitativos serán complementados con un análisis cualitativo de las observaciones formuladas por los expertos (Aracena et al., 2024). En función de los resultados obtenidos, se procederá de la siguiente manera: los ítems que tengan  $V \geq 0.80$  y sólida fundamentación teórica, se mantendrán; aquellos con V entre 0.70 y 0.79 se reformularán, incorporando las observaciones y sugerencias de los jueces expertos; y los ítems con  $V < 0.70$  o que presenten inconsistencias conceptuales se eliminarán.

### **Entrevistas cognitivas**

Como medida complementaria, se llevarán a cabo entrevistas cognitivas a un grupo de diez estudiantes que previamente hayan cumplido con los siguientes criterios de inclusión: encontrarse entre 2° y 4° básico, tener entre 7 a 10 años de edad, mostrar dominio del español y contar con el consentimiento y asentimiento informado correspondiente. Como criterio de exclusión, no se incluirá a los participantes con un

diagnóstico de Necesidades Educativas Especiales Permanentes (NEEP). Estas entrevistas permitirán verificar si los ítems son comprendidos de manera adecuada, lo que contribuirá a evaluar su equivalencia cultural desde la perspectiva de los propios participantes (Bellei, 2013; OECD, 2019).

Para ello, se utilizará un “Instrumento para la evaluación de Entrevista Cognitiva” (ver Anexo 2), la cual fue específicamente diseñada para este proyecto, con el fin de poder evaluar el pensamiento crítico en el razonamiento inferencial informal. El instrumento cuenta de una rúbrica, presentada en la Tabla 6, la cual mide cuatro criterios: (1) comprensión semántica y lingüística, (2) comprensión de la tarea y estructura, (3) comprensión conceptual y (4) engagement y pertinencia percibida.

**Tabla 6**

*Rúbrica de criterios para la evaluación de la entrevista cognitiva*

Criterio a evaluar	Indicadores observables	Escala Likert (1 - 5)	Comentarios cualitativos
Comprensión semántica y lingüística	<p>El niño no pregunta el significado de ninguna palabra.</p> <p>El niño puede parafrasear el enunciado usando sus propias palabras sin distorsionar el significado.</p> <p>Su expresión facial no muestra confusión por términos clave.</p> <p>No necesita que se le repita la pregunta de forma adicional.</p>		P. ej., “el estudiante frunce el ceño al leer la palabra inferencia.”
Comprensión de la tarea y estructura	Comprende que debe seleccionar una sola opción de respuesta.		P. ej., “requiere que se le explique

Comprende cómo marcar su respuesta sin ayuda.

nuevamente cómo marcar la respuesta.”

Entiende la instrucción general de la tarea.

El formato de selección múltiple (a-b-c-d) le es familiar y manejable.

Comprensión conceptual	Su explicación verbal del ítem se alinea con la habilidad de pensamiento crítico evaluada.  La alternativa elegida se ajusta con base en un razonamiento lógico, evidencial o comparativo.  No se distrae con elementos accesorios o irrelevantes del ítem.  Desacredita críticamente las alternativas no seleccionadas.	P. ej., “Se distrae con el dibujo de la botella y menciona colores.”
Engagement y pertinencia percibida	Mantiene la atención durante la lectura y resolución del ítem.  Reconoce que la situación presentada le es familiar o posible.  Muestra interés genuino por el contenido de la pregunta.  El tiempo que tarda en responder es apropiado (ni inmediato, ni excesivamente largo por distracción).	P. ej. “El estudiante demora 2:48 min en responder el ítem.”

---

Sugerencias de mejora

---

Cada categoría se operacionaliza mediante indicadores observables y una escala Likert de 1 a 5 puntos (Tabla 7), lo que permite integrar la información cualitativa y cuantitativa de manera coherente con la naturaleza de la investigación. Cada nivel de la escala Likert presenta un grado distinto de adecuación del ítem evaluado.

**Tabla 7**  
*Escala Likert para la evaluación del grado de adecuación de los ítems del instrumento*

1 punto	2 puntos	3 puntos	4 puntos	5 puntos
El ítem no cumple en absoluto con los indicadores. El ítem debe ser rediseñado completamente.	El ítem cumple deficientemente con los indicadores. El ítem requiere un ajuste estructural.	El ítem cumple de manera aceptable pero no óptima con los indicadores. El ítem requiere ajustes importantes.	El ítem cumple bien con los indicadores. La evidencia es mayoritaria y positiva. El ítem requiere menores ajustes.	El ítem cumple excelentemente. La evidencia es clara y fuerte en todos los aspectos (p. ej., los 4 indicadores son abordados de manera sobresaliente. El ítem no requiere modificaciones.

**Proceso de evaluación**

1. **Preguntas de sondeo:** Durante la resolución de cada ítem, el entrevistador formulará preguntas protocolarias como "¿Qué fue lo más difícil de entender?" o "¿Hay alguna palabra que no hayas comprendido?".
2. **Observación y registro:** Basándose en las respuestas verbales y conducta observada, el entrevistador completará la rúbrica asignando puntuaciones y registrando comentarios cualitativos.

3. **Interpretación de resultados:** Las puntuaciones se interpretarán según la guía de interpretación, donde 1-2 puntos indica problemas críticos, 3 puntos requiere ajustes necesarios, y 4-5 puntos refleja adecuación óptima.
4. **Generación de sugerencias:** En la sección final se documentarán las mejoras específicas recomendadas para cada ítem.

### **Protocolo para la realización de entrevistas cognitivas**

Durante las entrevistas, se emplearán técnicas como la de "pensar en voz alta", mediante la cual los estudiantes verbalizan sus pensamientos mientras responden, lo que permite acceder a los procesos cognitivos subyacentes. Asimismo, se aplicarán preguntas de sondeo o "probing" para clarificar la interpretación de las preguntas, identificar confusiones y comprender las razones detrás de sus elecciones (Smith-Castro y Molina-Delgado, 2011).

Todo este proceso se guiará mediante el *Manual de protocolo para la aplicación de la Entrevista Cognitiva* (ver [Anexo 3](#)). Este manual estandariza la evaluación basada en el "Instrumento para la Evaluación de Entrevista Cognitiva", y tiene como objetivo general garantizar la validez, confiabilidad y equidad del proceso de evaluación, minimizando los sesgos del entrevistador y obteniendo datos cualitativos robustos sobre la comprensión y el procesamiento mental de los estudiantes.

Sus objetivos específicos son:

- A. Estandarizar la interacción entre entrevistador y estudiante.
- B. Asegurar que las preguntas de sondeo se realicen de manera consistente.
- C. Guiar la observación y el registro de conductas observables.
- D. Uniformizar el proceso de calificación y generación de sugerencias de mejora para los ítems.

Entre los procedimientos específicos que contiene el manual, se destacan:

i. **Preparación pre-entrevista**, la cual define los requisitos del entorno, los materiales necesarios y el protocolo de presentación con el estudiante.

ii. **Aplicación de la entrevista**, la cual establece un flujo estructurado y preguntas de sondeo organizadas en 4 fases: comprensión semántica y lingüística, comprensión de la tarea y estructura, comprensión conceptual, y engagement y pertinencia percibida (Figura 7).

### Figura 7

#### *Preguntas de sondeo protocolarias por criterio*

##### **Fase 1: Evaluar la Comprensión Semántica y Lingüística**

Verificar que el estudiante entienda el significado literal de las palabras y el enunciado.

- "¿Hay alguna palabra o frase en esta pregunta que no hayas entendido?"
- "¿Entendiste la pregunta?¿necesitas que te la lea?"
- "¿Puedes explicarme con tus propias palabras de qué se trata el ejercicio?" (Busca una paráfrasis).

##### **Fase 2: Evaluar la Comprensión de la Tarea y Estructura**

Asegurarse de que el estudiante sabe QUÉ tiene que hacer y CÓMO hacerlo.

- "¿Qué es lo que tienes que hacer?"

##### **Fase 3: Evaluar la Comprensión Conceptual (Validez de Constructo)**

Profundizar en el razonamiento para confirmar que la respuesta se basó en la habilidad de pensamiento crítico que se pretende evaluar.

- "¿Qué te hizo pensar que esa era la respuesta correcta?"
- "¿Por qué descartaste las otras opciones?"

##### **Fase 4: Evaluar el Engagement y Pertinencia Percibida**

Capturar la experiencia subjetiva del estudiante y su nivel de conexión con la tarea.

- "¿Qué te pareció el tema de la tarea?"
- La situación que se presenta o algo parecido ¿la has vivido en algún momento?

NO: ¿crees que es posible que pase algo así en la vida real?

SI: ¿en qué momento la viviste?

iii. **Técnicas de observación y registro**, la cual orienta al entrevistador a captar las señales no verbales y en la redacción de comentarios cualitativos basados en evidencias.

iv. **Calificación post-entrevista**, la cual describe el procedimiento para la asignación de puntuaciones Likert (1-5), así como la generación de sugerencias de mejora y la interpretación de resultados.

v. **Principios éticos**, que indica mantener neutralidad, confidencialidad y un trato respetuoso.

Las entrevistas serán grabadas en audio con el consentimiento previo de los estudiantes y sus tutores, y se tomarán notas detalladas, para permitir un análisis posterior riguroso. Este análisis evaluará los datos en función de los criterios previamente definidos, con el objetivo de identificar patrones de comprensión errónea, ambigüedades en la redacción de los ítems o problemas relacionados con la equivalencia cultural desde la perspectiva de los propios participantes. De esta manera, se obtendrán datos que orienten la mejora de la claridad, pertinencia y adecuación cultural del instrumento, contribuyendo a su validez basada en el proceso de respuesta (Caicedo y Zalazar, 2018).

### **Proceso para el análisis de datos obtenidos de las entrevistas cognitivas**

Una vez finalizadas las entrevistas y completadas las rúbricas, se llevará a cabo una sesión de análisis colaborativo entre las investigadoras (entrevistadoras) y el profesor guía. El primer paso consistirá en consolidar los datos cuantitativos y cualitativos en una hoja de cálculo. Esta hoja estará estructurada con las siguientes columnas: ítem, entrevistado, alternativa de respuesta seleccionada, los criterios evaluados y la edad de los participantes. En ella se registrará, para cada ítem y estudiante, la puntuación Likert (de 1 a 5) obtenida en cada uno de los cuatro criterios del Instrumento para la Evaluación de la Entrevista Cognitiva (Figura 8).

## Figura 8

*Ejemplo de hoja de cálculo con datos cuantitativos del ítem 1 y 2*

Ítem	Entrevistado	Alternativa	Comprensión semántica y lingüística	Comprensión de la tarea	Comprensión conceptual	Engagement y pertinencia percibida	Edad
1	M	c	5	5	5	5	10
1	A	c	5	5	5	5	8
2	M	b	5	5	5	5	10
2	A	b	5	5	5	5	8

Posteriormente, para cada ítem se calculará la puntuación promedio por criterios, a partir de los resultados obtenidos de los entrevistados. La interpretación de estos promedios y la consiguiente decisión sobre cada ítem se registrarán por el siguiente criterio preestablecido anteriormente en el “Instrumento para la evaluación de Entrevista Cognitiva”:

- **Puntuación promedio de 5:** Se considera que el ítem es óptimo en ese criterio, por lo que no requiere modificaciones.
- **Puntuación promedio de 4:** Indica la necesidad de ajustes menores para mejorar la claridad o precisión del ítem.
- **Puntuación promedio de 3 o inferior:** Señala problemas sustanciales que requieren ajustes mayores o una reformulación importante.

Los resultados cuantitativos de las entrevistas cognitivas se complementarán con el análisis cualitativo de los comentarios de los entrevistadores, los cuales registran las verbalizaciones de los estudiantes, sus dudas sobre palabras y sus razonamientos expresados durante el sondeo; de las razones detrás de las respuestas incorrectas; y de las observaciones de conducta. De manera colaborativa, el equipo determinará y documentará las modificaciones

### **3.5 CONSIDERACIONES ÉTICAS**

La presente investigación se desarrollará bajo principios éticos específicos para estudios con población infantil, respetando su dignidad, autonomía progresiva y derecho a participar de forma activa en decisiones que les afecten. Tal como plantea el Centro de Investigaciones Innocenti (2013) en la Carta Internacional para la Investigación Ética de la Niñez, los investigadores son responsables de promover prácticas respetuosas, justas y seguras, que protejan a los niños y niñas durante todo el proceso de investigación.

Se solicitará el consentimiento informado a madres, padres o apoderados, y de manera complementaria, se gestionará el asentimiento de los estudiantes participantes, asegurando que comprendan, en un lenguaje acorde a su edad, los objetivos, procedimientos y alcances del estudio. Estos documentos serán entregados para la realización de las entrevistas cognitivas, de modo que se garantice la participación libre, informada y voluntaria en la investigación.

Durante todo el proceso se garantizará la confidencialidad y el anonimato de la información recopilada, resguardando la identidad de los participantes y el uso ético de los datos obtenidos. Finalmente, se actuará con especial cuidado para no causar daño físico, emocional y cognitivo a los estudiantes, priorizando su bienestar por sobre cualquier objetivo investigativo.

#### IV. ANÁLISIS Y RESULTADOS

A continuación, se presentan los resultados obtenidos correspondientes a las dos etapas del proceso de validación del instrumento diseñado, correspondientes a las entrevistas cognitivas a estudiantes y las evaluaciones de jueces expertos.

##### 4.1 RESULTADOS DERIVADOS DE LAS ENTREVISTAS COGNITIVAS

Las entrevistas cognitivas se aplicaron a un grupo de diez estudiantes. El análisis se realizó con base en cuatro criterios: comprensión semántica y lingüística, comprensión de la tarea, comprensión conceptual, y engagement y pertinencia percibida. Cada uno de los criterios fue evaluado en una escala Likert del 1 al 5 (Tabla 7).

Para la interpretación de los resultados se consideró que: una puntuación promedio de 5 indica que el ítem es óptimo en el criterio y no requiere modificaciones; una puntuación promedio de 4 señala la necesidad de ajustes menores para mejorar la claridad o precisión; mientras que una puntuación promedio de 3 o inferior evidencia problemas sustanciales que requieren ajustes mayores o una reformulación importante.

Los datos fueron tabulados a una hoja de cálculo, donde se determinaron los puntajes promedio por criterio para cada ítem (Tabla 8).

**Tabla 8**  
*Valoraciones de entrevistas cognitivas según criterios de evaluación*

	Comprensión semántica y lingüística	Comprensión de la tarea y estructura	Comprensión conceptual	Engagement y pertinencia percibida
Ítem 1	5,0	5,0	5,0	5,0
Ítem 2	5,0	5,0	5,0	5,0
Ítem 3	3,0	5,0	3,0	5,0
Ítem 4	3,5	5,0	3,5	4,5
Ítem 5	4,7	5,0	4,7	4,7
Ítem 6	5,0	5,0	4,5	5,0
Ítem 7	4,5	5,0	4,5	5,0

Ítem 8	5,0	5,0	5,0	5,0
Ítem 9	3,5	4,5	4,0	5,0
Ítem 10	4,5	5,0	4,0	5,0
Ítem 11	5,0	4,5	5,0	5,0
Ítem 12	5,0	5,0	5,0	5,0
Ítem 13	3,5	5,0	4,5	4,5
Ítem 14	5,0	5,0	4,0	5,0
Ítem 15	4,5	5,0	5,0	5,0
Ítem 16	4,0	5,0	4,0	5,0

*Nota.* Los puntajes destacados en color rojo representan los puntajes críticos obtenidos en las entrevistas cognitivas de acuerdo a la escala Likert.

A partir de los resultados presentados en la Tabla 8, se observa un desempeño general favorable en todos los criterios evaluados. De los 16 ítems presentados, el 75% de ellos presentó puntajes entre 4 y 5 en todos los criterios, lo que indica que cumplieron bien o excelentemente con los indicadores evaluados. Sin embargo, se identificaron cuatro ítems con puntuaciones inferiores a 4 en ciertos criterios, lo que evidenció que cumplieron de manera aceptable pero no óptima con los indicadores.

En función de los resultados obtenidos, se llevó a cabo un análisis para realizar los ajustes necesarios por criterio, los cuales se detallan a continuación.

#### 4.1.1 Comprensión semántica y lingüística

Este criterio evalúa la claridad, precisión y adecuación del lenguaje utilizado en los ítems, considerando si los estudiantes comprenden correctamente las palabras y expresiones, si son capaces de parafrasear los enunciados sin alterar su significado y si no presentan signos de confusión.

Los resultados obtenidos evidencian que el 75% de los ítems alcanzó puntajes entre 4 y 5. De acuerdo a la escala Likert utilizada, aquellos ítems que tuvieron puntaje 5 cumplieron excelentemente con los indicadores propuestos, sin requerir modificaciones, mientras que aquellos que obtuvieron 4, 4,5 y 4,7 cumplieron bien con los indicadores, requiriendo ajustes menores para mejorar la comprensión del ítem. Sin embargo, el 15% restante, correspondiente a los ítems 3, 4, 9 y 13 obtuvieron puntajes entre 3 y 3,5, lo que

sugiere que los indicadores se cumplieron de manera aceptable pero no óptima, señalando problemas sustanciales que requieren ajustes.

En el caso del ítem 3 (Figura 9) que obtuvo una puntuación promedio de 3, el análisis de las entrevistas cognitivas reveló que las dificultades se debieron a un referente cultural poco familiar para los estudiantes, siendo este el canal de televisión “Disney”.

### Figura 9

*Versión original del ítem 3 utilizado en el “Instrumento para la evaluación de Entrevista Cognitiva”*

<p>Camilo quiere saber qué canal de televisión prefieren ver en su curso. Preguntó a 2 de ellos y concluyó que Disney es el preferido.</p>
<p>¿Es suficiente lo que hizo Camilo para responder?</p> <ul style="list-style-type: none"><li>a) No, debería preguntar a más compañeros.</li><li>b) Sí, porque a dos les gusta.</li><li>c) Sí, Disney es el mejor canal.</li><li>d) No, debería preguntar al profesor.</li></ul>

Los comentarios cualitativos realizados en el proceso de las entrevistas cognitivas revelan que uno de los estudiantes manifestó desconocer el término “Disney”, mientras que otro estudiante comprendió de manera errónea el ítem. Además, ambos tuvieron problemas para parafrasear correctamente el enunciado.

Ante esa evidencia, se optó por sustituir el término “Disney” por “Mr. Best”, que es un nombre ficticio para hacer referencia a un creador de contenido real y conocido por los estudiantes, reemplazando también “canal de televisión” por “youtuber”. Estos cambios favorecen una mejor comprensión, ya que introduce un referente cultural familiar y coherente con el entorno cotidiano del grupo evaluado. De esta manera, se mejora la comprensión semántica y lingüística del ítem al eliminar posibles ambigüedades asociadas a términos desconocidos y se potencia la adecuación lingüística, al emplear un vocabulario más acorde al nivel cognitivo y sociocultural de los participantes.

En esta misma línea, se encuentra el ítem 4 (Figura 10), que obtuvo un puntaje de 3,5. Durante las entrevistas cognitivas, se observó que la palabra “jugo” podía ser confundida con la palabra “juego”. Este problema fue evidente en una de las entrevistas, donde el estudiante mostró confusión hacia ambos términos y, al momento de parafrasear, expresó que debía “ver cuál es el juego que menos juegan”, distorsionando así el sentido original del ítem.

### **Figura 10**

*Versión original del ítem 4 utilizado en el instrumento “Instrumento para la evaluación de Entrevista Cognitiva”*

David hizo una encuesta en su curso para saber cuál es el jugo preferido del 1°A. Preguntó a sus 3 mejores compañeros y concluyó que el jugo de piña es el menos popular del curso.
¿Podemos confiar en lo que dice David?
a) No, porque les preguntó a pocos compañeros
b) Sí, porque solo uno voto por piña.
c) Sí, porque a sus amigos les gusta naranja y manzana.
d) No, porque el jugo de piña es saludable.

En general, los participantes no manifestaron confusión frente a los términos clave. Sin embargo, mostraron dificultad para parafrasear correctamente el enunciado completo, lo que sugiere que la ambigüedad semántica del término “jugo” interfirió en la comprensión semántica y lingüística.

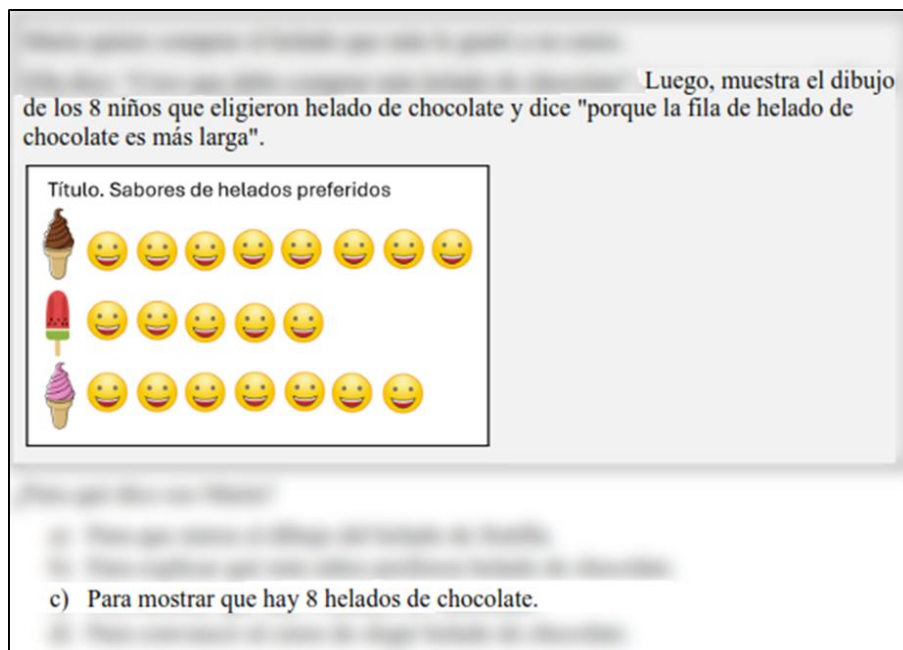
Con el fin de mejorar esto, se decidió sustituir la palabra “jugo” por “bebida”, que es un término menos ambiguo al momento de leerlo. Este ajuste permite asegurar una interpretación más precisa, fortaleciendo la comprensión semántica y lingüística del ítem.

En el ítem 9 (Figura 11), se obtuvo un puntaje promedio de 3,5 puntos. El análisis de las entrevistas cognitivas evidenció que la expresión “helado de leche” no resultaba suficientemente clara para los estudiantes, generando confusión respecto a su significado

y relación con la imagen del ítem. En ambas entrevistas, los participantes asociaron dicho término con “helado de vainilla”, lo que indica una interpretación distinta a la esperada.

### Figura 11

*Versión original del ítem 9 utilizando en el “Instrumento para la evaluación de Entrevista Cognitiva”*



Durante las entrevistas, los niños no preguntaron el significado de las palabras ni mostraron expresiones faciales de confusión; sin embargo, uno de ellos necesitó que se le repitiera la pregunta, y el otro presentó dificultades al parafrasear correctamente el enunciado, lo que derivó en una comprensión parcial del ítem. Además, se identificó confusión en la alternativa c), relacionada con la presencia del distractor numérico “8”, ya que uno de los estudiantes eligió inicialmente esa opción basándose en el número de caritas felices que observaba en el gráfico.

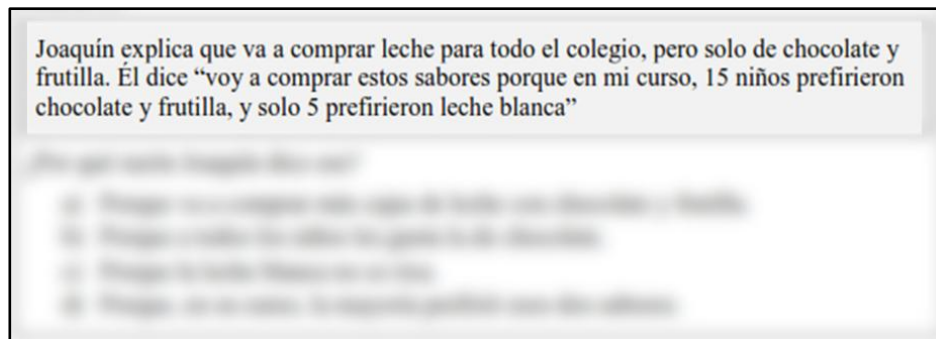
A partir de esta evidencia, se decidió reemplazar la expresión “helado de leche” por “helado de chocolate”, utilizando un referente más familiar para los estudiantes. Asimismo, se eliminó el distractor “8” con el fin de evitar confusiones y mejorar la

interpretación de las alternativas, contribuyendo a optimizar la comprensión semántica y lingüística del ítem.

Por último, el ítem 13 (Figura 12) obtuvo una puntuación de 3,5. El análisis de las entrevistas cognitivas reveló que el principal problema se encontraba en la redacción redundante del enunciado, lo que dificultaba la comprensión semántica y lingüística del ítem.

### **Figura 12**

*Versión original del ítem 13 utilizando en el “Instrumento para la evaluación de Entrevista Cognitiva”*



En las entrevistas cognitivas, ninguno de los estudiantes manifestó confusión frente a los términos clave ni solicitó aclaraciones sobre el significado de las palabras. No obstante, ambos estudiantes presentaron dificultades para parafrasear el enunciado con sus propias palabras sin alterar su sentido, lo que evidencia una comprensión parcial del ítem. Además, uno de ellos requirió que se le repitiera la pregunta en varias ocasiones y no comprendió el significado de la alternativa b), señalando que, aunque entendía todas las palabras y frases empleadas, la redacción del enunciado le resultó difícil de comprender.

Ante esa evidencia, se determinó que el problema radicaba en la extensión y redundancia del texto. Por ello, se procedió a simplificar el enunciado, eliminando expresiones repetitivas y manteniendo únicamente aquellos elementos necesarios para

transmitir la idea principal de manera clara y directa. Esta modificación busca favorecer la fluidez lectora, mejorando así la comprensión semántica y lingüística.

#### 4.1.2 Comprensión de la tarea y estructura

Este criterio evalúa la capacidad de los estudiantes para comprender la instrucción general de la tarea, reconocer que debían seleccionar una opción como respuesta y marcarla correctamente sin necesitar ayuda adicional. Además, considera el nivel de familiaridad de los estudiantes con el formato de selección múltiple (a, b, c y d).

De los 16 ítems propuestos, el 100% muestra un buen desempeño. El 87,5% de ellos obtuvo 5 puntos, lo que indica que cumplen excelentemente con los indicadores evaluados, sin necesitar modificaciones. Por otro lado, los ítems 9 y 11, correspondientes al 12,5% restante, obtuvieron 4,5 puntos, lo que indica que cumplen bien con los indicadores, requiriendo sólo cambios menores.

En conjunto, estos resultados sugieren que los ítems presentan una estructura precisa, con instrucciones claras y sin ambigüedades que pudieran interferir en la interpretación o ejecución de la tarea. Dado que no se detectaron dificultades significativas ni errores de comprensión en este criterio, no fue necesario realizar modificaciones en los ítems evaluados.

Los comentarios cualitativos realizados durante las entrevistas cognitivas respaldan estos hallazgos, ya que los estudiantes manifestaron comprender adecuadamente las instrucciones generales y el formato de selección múltiple. En la mayoría de los casos, señalaron que las consignas son claras y explícitas, sin requerir apoyo adicional. No obstante, en los ítems 9 y 11 los estudiantes preguntaron cómo debían marcar la alternativa correcta.

Estas dificultades podrían atribuirse al hecho de que, en ambos casos, se trataba del primer ítem presentado durante la aplicación, momento en el cual los estudiantes aún se encontraban familiarizándose con el formato y la dinámica de la tarea. Una vez

comprendida la modalidad de respuesta, no se registraron inconvenientes en los ítems posteriores, lo que refuerza la interpretación de que las dudas observadas fueron iniciales y de carácter procedimental, más que relacionadas con la comprensión de la tarea y estructura.

#### 4.1.3 Comprensión conceptual

Este criterio evalúa la capacidad de los estudiantes en la comprensión conceptual de los estudiantes en relación con la habilidad de pensamiento crítico que se busca medir. Considera si los estudiantes explican correctamente el sentido del ítem, justifican su respuesta mediante un razonamiento lógico o basado en evidencias, y logran diferenciar la información relevante de la irrelevante. Además, se observa si son capaces de rechazar críticamente las alternativas incorrectas, lo que refleja una comprensión adecuada del contenido evaluado.

Los resultados evidencian que, de los 16 ítems evaluados, el 37,5% obtuvo una puntuación de 5, lo que indica que cumplen excelentemente con los indicadores establecidos y no requieren modificaciones. Por otra parte, el 50% alcanzó puntajes de 4, 4,5 y 4,7, lo que refleja un buen cumplimiento de los indicadores, aunque con necesidad de realizar ajustes menores para optimizar la comprensión conceptual. Finalmente, el 12,5% de los ítems presentó puntuaciones de 3 y 3,5, lo que sugiere un cumplimiento aceptable, pero no óptimo, por lo que requieren modificaciones mayores.

El ítem 3 (Figura 9) obtuvo una puntuación de 3. El análisis de las entrevistas cognitivas evidenció que la dificultad principal se relacionó con la presencia del término “Disney” tanto en el enunciado como en la alternativa c). Este elemento actuó como un distractor, ya que su repetición generaba un sesgo de familiaridad que llevaba a los estudiantes a seleccionar esa opción sin basarse en un razonamiento conceptual, sino en el reconocimiento del término.

En una de las entrevistas, la explicación del estudiante se alineó con la habilidad del pensamiento crítico evaluada, y su elección de respuesta se basó en un razonamiento

comparativo. Sin embargo, se distrajo con elementos accesorios del ítem y no fue capaz de desacreditar críticamente las alternativas no seleccionadas. Además, descartó la alternativa a) por motivos ajenos al contenido conceptual, argumento que se puede acabar el tiempo de la clase si les pregunta a más compañeros, y terminó eligiendo la alternativa que contenía la palabra que desconocía.

En otra entrevista, la explicación del estudiante se alineó con la habilidad del pensamiento crítico evaluada, y la alternativa seleccionada no se sustentó en un razonamiento lógico o evidencial. Aunque no se distrajo con elementos accesorios y desacreditó críticamente las alternativas no seleccionadas, manifestó la necesidad de releer el ítem y centró su análisis en dos personas, sin considerar al curso completo.

A partir de estas observaciones, se determinó que la presencia del término “Disney” interfería tanto en la interpretación como en la validez conceptual del ítem. Por ello, se decidió sustituirlo por “Mr. Best”, un referente cultural más familiar para los estudiantes, y eliminar su repetición en las alternativas de respuesta.

Por último, el ítem 4 (Figura 10) obtuvo una puntuación de 3,5. En este caso, la confusión asociada con la palabra “jugo” dificultó la interpretación del enunciado, ya que algunos estudiantes lo asociaron erróneamente con el término “juego”. Esta confusión léxica alteró el sentido del problema, generando un cambio en la situación planteada y afectando la comprensión conceptual del ítem. Al interpretar el ítem desde una lógica distinta a la prevista, los estudiantes desviaron su razonamiento hacia aspectos no vinculados con la habilidad de pensamiento crítico que se buscaba evaluar. En consecuencia, el ítem no garantizaba una interpretación conceptual uniforme entre los participantes, lo que compromete su validez de contenido, al no asegurar que todos los estudiantes comprendan el mismo concepto o proceso cognitivo evaluado.

En una de las entrevistas, la explicación verbal del estudiante se alineó con la habilidad de pensamiento crítico evaluada, y la alternativa elegida se sustentó en un razonamiento lógico, evidencial o comparativo. Sin embargo, se distrajo con elementos

accesorios o irrelevantes del ítem y no logró desacreditar críticamente las alternativas no seleccionadas. Además, expresó que no comprendía completamente el enunciado debido a la falta de información sobre cuántas personas habían votado, lo que lo llevó a releer el ejercicio para intentar resolver la confusión.

En la segunda entrevista, la explicación verbal del estudiante también se alineó con la habilidad de pensamiento crítico evaluada, y la elección de la alternativa se basó en un razonamiento lógico o evidencial. Al igual que en el caso anterior, se distrajo con elementos accesorios del ítem, aunque en esta ocasión sí fue capaz de desacreditar críticamente las alternativas no seleccionadas.

A partir de estas observaciones, se concluyó que la ambigüedad generada por el término “jugo” no solo afectaba la comprensión semántica del enunciado, sino también la comprensión conceptual, al inducir una lectura errónea del contexto y modificar el tipo de razonamiento implicado. Este tipo de error compromete la validez de contenido del ítem, ya que impide que mida de manera precisa la habilidad de pensamiento crítico que pretende evaluar. Por ello, se decidió sustituir “jugo” por “bebida”, con el fin de garantizar una interpretación unívoca, coherente con el constructo evaluado, y fortalecer la consistencia conceptual del ítem.

#### 4.1.4 Engagement y pertinencia percibida

Este criterio analiza el nivel de interés que los estudiantes demostraron frente a las situaciones planteadas en los ítems, así como el grado en que dichas situaciones resultaron comprensibles, cercanas y coherentes con su experiencia cotidiana. También considera la capacidad de los estudiantes para mantener la atención durante la lectura y responder con un tiempo adecuado, sin evidenciar distracción o desinterés.

El 100% de los ítems evidencian un buen *engagement* y pertinencia. De ellos, el 81,25% obtuvo la puntuación máxima (5), reflejando un cumplimiento excelente de los indicadores, lo que implica que no se deben realizar cambios en aquellos ítems. Mientras

tanto, el 18,75% obtuvo puntuaciones entre 4,5 y 4,7, lo que muestra que se cumplió bien con los indicadores, siendo necesario realizar pequeños ajustes.

Además, a través de los comentarios cualitativos realizados en las entrevistas cognitivas, se evidenció que el tiempo promedio de respuesta fue de 2 minutos con 20 segundos, lo que es un tiempo adecuado, sin indicios de distracción ni de respuestas apresuradas. Aquellos ítems que presentaron tiempos más largos, como lo fue el caso del ítem 13 con un tiempo promedio de 4 minutos con 34 segundos, se asociaron a las dificultades presentadas en los criterios anteriores, las cuales fueron posteriormente corregidas.

En conjunto, los ítems presentaron contextos atractivos, significativos y acordes al entorno de los estudiantes, lo que se refleja en los altos puntajes de engagement y pertinencia y en la mantención de la atención durante la tarea.

## **4.2 RESULTADOS DERIVADOS DE LA EVALUACIÓN DE JUECES EXPERTOS**

Este apartado presenta los resultados obtenidos a partir de la evaluación realizada por siete jueces expertos. La valoración se organizó en seis criterios: claridad, pertinencia, contextualización para el razonamiento inferencial informal, relevancia de las alternativas y cobertura del pensamiento crítico por ítem. Cada criterio fue evaluado en una escala Likert de 1 a 5 puntos (Tabla 7).

La evaluación de jueces expertos se llevó a cabo mediante dos procesos sucesivos. En la primera revisión, los resultados evidenciaron que el instrumento incluía ítems adecuados en términos generales, pero también detectaron ítems críticos (Tabla 9), lo que implica que obtuvieron puntuaciones menores de tres, referentes a la escala Likert, en uno o más criterios. Estas observaciones mostraron dificultades en todos los criterios evaluados, aunque concentradas en ciertos ítems específicos. En varios casos se identificó una redacción ambigua o una evaluación insuficiente de la sub-habilidad comprendida.

También se detectaron distractores que no cumplían adecuadamente con su función, por ser rápidamente descartables, poco diferenciadores o no tener coherencia con la situación planteada en los ítems. En conjunto, estas evaluaciones reflejaron la necesidad de realizar ajustes que permitieran mejorar la calidad del instrumento.

**Tabla 9**  
*Puntuaciones obtenidas en ítems críticos en la primera evaluación de jueces expertos*

Jueces	Ítem	Claridad	Pertinencia	Contextualización para el razonamiento inferencial informal	Relevancia de las alternativas	Cobertura del pensamiento crítico
A	5	4	4	4	4	4
C	5	4	4	4	4	4
J	5	2	2	2	2	NA
M	5	2	4	2	4	3
N	5	4	4	4	4	4
P	5	3	4	3	4	4
R	5	2	3	2	2	4
A	11	4	4	4	4	4
C	11	3	3	3	3	4
J	11	3	2	4	3	NA
M	11	2	2	2	3	3
N	11	3	4	3	4	4
P	11	3	4	3	4	4
R	11	3	2	3	4	2
A	13	4	4	4	4	4
C	13	4	3	4	4	4
J	13	4	4	4	4	NA
M	13	3	2	2	3	3
N	13	3	4	4	4	4
P	13	3	3	3	3	4
R	13	4	1	1	2	1
A	15	4	4	4	4	4
C	15	4	4	4	4	4

J	15	4	2	2	1	NA
M	15	3	2	3	2	1
N	15	4	4	4	4	4
P	15	3	4	4	3	3
R	15	3	4	3	NA	NA
A	16	4	4	4	4	4
C	16	4	4	4	3	4
J	16	2	2	2	2	NA
M	16	3	2	3	2	1
N	16	4	4	4	4	4
P	16	4	4	4	4	4
R	16	3	4	4	3	2

*Nota.* Las celdas en rojo indican puntuaciones bajas (1-2) o datos faltantes, que requieren revisión.

La identificación de los ítems críticos se sustenta en cuatro criterios complementarios:

**Criterio 1. Acumulación de calificaciones bajas:** El primer criterio para identificar los ítems críticos es la cantidad total de calificaciones bajas recibidas por parte de los jueces en los distintos criterios evaluados (claridad, pertinencia, razonamiento inferencial informal, relevancia de las alternativas y pensamiento crítico por ítem). Los ítems catalogados como críticos concentraron entre 7 y 10 observaciones negativas, lo que evidencia la presencia de debilidades distribuidas en múltiples dimensiones del instrumento. Este patrón indica que los problemas del ítem no se limitan a un aspecto aislado, sino que se manifiestan de manera simultánea en varios criterios. En contraste, los ítems que, si bien requieren ajustes, no fueron clasificados como críticos, presentan un máximo de 7 calificaciones bajas (sin cumplir los otros criterios para determinar si es un ítem crítico) o un mínimo de 1 calificación baja, lo que sugiere un nivel de dificultad menor y más focalizado.

**Criterio 2. Coincidencia entre jueces en la asignación de calificaciones bajas:** El segundo criterio determina que un ítem es considerado crítico cuando tres o más jueces coincidieron en asignarle una calificación baja en alguno de los criterios evaluados. Este

patrón de convergencia indica que la debilidad del ítem no corresponde a una observación aislada o subjetiva, sino a un problema reconocido por múltiples evaluadores. La recurrencia de puntajes bajos entre distintos jueces revela dificultades consistentes del ítem, lo que justifica su clasificación como elemento prioritario para ser revisado.

**Criterio 3. Presencia de calificaciones bajas en múltiples criterios por parte de distintos jueces:** Asimismo, se considera que un ítem es crítico cuando dos jueces o más le asignan puntajes bajos en al menos dos criterios distintos. Este patrón refleja no solo coincidencia entre evaluadores, sino también una amplitud de las dificultades detectadas, ya que las observaciones negativas se distribuyen en más de una dimensión del instrumento. La presencia de debilidades simultáneas en diferentes criterios (como claridad, pertinencia o razonamiento inferencial informal) evidencia problemas más complejos en el diseño del ítem y fundamenta la necesidad de realizar ajustes profundos.

**Criterio 4. Formación de grupos de debilidades dentro de los criterios:** El cuarto criterio de identificación consideró la formación de grupos de calificaciones bajas dentro de un mismo criterio, es decir, la repetición de observaciones negativas por parte de distintos jueces en una misma dimensión del ítem. Los ítems críticos se caracterizaron por presentar entre dos y tres grupos de este tipo, lo que revela patrones de debilidad consistentes y recurrentes en distintas áreas del instrumento. Por ejemplo, un ítem puede acumular dos evaluaciones negativas en pertinencia, dos en razonamiento inferencial informal, tres en relevancia de las alternativas, conformando así tres grupos claramente identificables. En comparación, los ítems no críticos tienden a mostrar uno o dos grupos (sin cumplir los otros criterios para determinar si es un ítem crítico), o bien ninguno, lo que indica que las observaciones negativas son más aisladas y menos coincidentes entre los jueces.

Con base en estas observaciones, se procedió a reformular los ítems que presentaron mayores dificultades, correspondientes al 5, 11, 13, 15 y 16. Los ajustes incluyeron la simplificación del lenguaje, la reestructuración de enunciados extensos, la

modificación de alternativas, etc. Una vez realizadas estas modificaciones, se solicitó una segunda revisión a los mismos jueces expertos.

Los resultados de la segunda evaluación (Tabla 10) muestran una mejora general en la valoración de los ítems. En comparación con la primera versión, se observa un aumento consistente en la puntuación de criterios como claridad y pertinencia, junto con una mayor estabilidad en las revisiones de los jueces. De acuerdo a los criterios de identificación definidos previamente, ninguno de los ítems se considera crítico en esta segunda revisión, lo que indica que los ajustes realizados permitieron resolver varias de las dificultades detectadas inicialmente.

**Tabla 10**

*Puntuaciones obtenidas en ítems críticos en la segunda evaluación de jueces expertos*

Jueces	Ítem	Claridad	Pertinencia	Contextualización para el razonamiento inferencial informal	Relevancia de las alternativas	Cobertura del pensamiento crítico
A	5	4	4	4	4	4
C	5	4	4	4	4	4
J	5	4	4	4	4	3
M	5	4	4	3	4	3
N	5	4	4	4	4	4
P	5	3	4	3	4	4
R	5	4	3	4	4	4
A	11	4	4	4	4	4
C	11	3	3	3	3	4
J	11	4	3	4	4	4
M	11	4	4	4	4	4
N	11	3	4	3	4	4
P	11	3	4	3	4	4
R	11	3	3	3	4	3
A	13	4	4	4	4	4
C	13	4	3	4	4	4

J	13	4	4	4	4	2
M	13	3	2	2	3	3
N	13	3	4	4	4	4
P	13	3	4	4	3	4
R	13	4	3	3	3	3
A	15	4	4	4	4	4
C	15	4	4	4	4	4
J	15	4	4	4	4	3
M	15	3	2	3	2	1
N	15	4	4	4	4	4
P	15	3	4	4	3	3
R	15	3	4	3	4	4
A	16	4	4	4	4	4
C	16	4	4	4	3	4
J	16	4	4	4	4	4
M	16	3	3	3	3	3
N	16	4	4	4	4	4
P	16	4	4	4	4	4
R	16	3	4	4	3	4

*Nota.* Aquellos puntajes críticos que se incrementaron con respecto a la evaluación anterior fueron resaltados en color amarillo, mientras que aquellos en verde representan aquellos puntajes que no pudieron ser reevaluados por los jueces.

#### 4.2.1 Síntesis por criterio

A continuación, se presenta una síntesis de los resultados obtenidos en las evaluaciones de los jueces expertos para cada uno de los criterios considerados.

##### **Claridad**

En este criterio se observaron diferencias importantes entre la primera y la segunda evaluación, lo que permite identificar con precisión el efecto de los ajustes realizados en los enunciados. En la primera revisión (Tabla 11), ocho ítems presentaron al menos una valoración inferior a 3 por parte de los jueces, lo que evidenciaba problemas asociados a enunciados extensos o palabras ambiguas. Además, uno de los jueces decidió no calificar

este criterio en uno de los ítems. Tras la reformulación del instrumento, sólo cuatro ítems muestran puntuaciones inferiores a 3, observándose una mayor consistencia entre evaluadores, lo que indica una mejora en la claridad de la mayoría de los ítems.

**Tabla 11**

*Puntuaciones obtenidas de ítems críticos y en revisión para el criterio de “Claridad” en la primera y segunda evaluación de jueces expertos*

Jueces	Ítem	Claridad	
		Primera evaluación	Segunda evaluación
R	2	2	2
J	3	NA	4
J	5	2	4
M	5	2	4
R	5	2	4
M	6	2	3
P	6	2	2
J	9	3	4
M	9	2	3
J	11	3	4
M	11	2	4
M	12	2	2
M	14	2	2
J	16	2	4

*Nota.* El rojo indica ítems con puntajes críticos y el rosado ítems que requieren revisión. El amarillo corresponde a aquellos puntajes que subieron. Las celdas sin color corresponden a puntuaciones que se mantuvieron sin cambios.

El análisis por ítem refleja con mayor detalle estas mejoras. En la primera evaluación, el ítem 2, 9, 11, 12, 14 y 16 recibieron una valoración de 2 puntos, mientras que el ítem 6 obtuvo dos valores de 2 puntos y el ítem 5, tres valoraciones de 2 puntos. Por otra parte, el ítem 3 no recibió valoración por parte de uno de los jueces. Estos datos muestran que, en su versión inicial, la claridad era un problema relevante en el instrumento.

En la segunda evaluación (Tabla 11), la cantidad de puntuaciones inferiores a 3 disminuye de manera evidente. Sólo los ítems 2, 6, 12 y 14 mantienen una valoración de 2 puntos por parte de algún juez, lo que indica que, si bien persisten ciertas dificultades

puntuales, éstas son mucho más acotadas que en la versión inicial. En el resto de los ítems desaparecen las valoraciones inferiores a 3 puntos, lo que demuestra una mejora notable tras los ajustes realizados.

Además, varios ítems que en la primera evaluación recibían valoraciones de 3 puntos muestran puntuaciones más altas y una menor variabilidad en la segunda evaluación. Esto muestra que las modificaciones permitieron corregir errores problemáticos y también optimizar aspectos de redacción que se encontraban en un nivel intermedio.

Además, varios ítems que en la primera evaluación recibían valoraciones de 3 puntos muestran puntuaciones más altas y una menor variabilidad en la segunda evaluación. Esto muestra que las modificaciones permitieron corregir errores problemáticos y también optimizar aspectos de redacción que se encontraban en un nivel intermedio.

### **Pertinencia**

En el criterio de “Pertinencia”, se observaron cambios entre la primera y segunda evaluación del instrumento. En la revisión inicial (Tabla 12), doce ítems presentaron al menos una valoración inferior a 3 por parte de los jueces, lo que significa que presentaban dificultades en la alineación con la habilidad y sub-habilidad que está midiendo, o la presencia de elementos distractores.

**Tabla 12**

*Puntuaciones obtenidas de ítems críticos y en revisión para el criterio de “Pertinencia” en la primera y segunda evaluación de jueces expertos*

Jueces	Ítem	Pertinencia	
		Primera evaluación	Segunda evaluación
J	1	2	4
J	3	NA	4
M	3	2	2
M	4	2	2
J	5	2	4

M	6	2	2
M	9	2	3
M	10	2	2
J	11	2	3
M	11	2	4
R	11	2	3
M	12	2	2
M	13	2	2
R	13	1	3
J	15	2	4
M	15	2	2
J	16	2	4
M	16	2	3

*Nota.* El rojo indica ítems con puntajes críticos y el rosado ítems que requieren revisión. El amarillo corresponde a aquellos puntajes que subieron. Las celdas sin color corresponden a puntuaciones que se mantuvieron sin cambios.

El análisis detallado por ítem revela que, en la primera evaluación, los doce ítems presentes en la Tabla 12 recibieron al menos una valoración de 2 puntos, mientras que el ítem 13 obtuvo una clasificación de 1 punto y el ítem 3 no recibió valoración por parte de uno de los jueces, indicando problemas graves de pertinencia.

Tras la reformulación de los ítems que requerían mejora, la segunda evaluación muestra una mejora en este criterio. De los doce ítems que presentaban problemas, sólo siete mantienen valoraciones de 2 puntos por parte de un juez. El resto de los ítems que en la primera evaluación fueron criticados, lograron puntuaciones de 3 o 4 puntos, siendo efectivos con la alineación de los ítems con las sub-habilidades correspondientes.

Los resultados de la segunda evaluación (Tabla 12) confirman los valores cuantitativos acerca de las mejoras descritas anteriormente. La mayoría de los ítems que inicialmente eran críticos o requerían mejoras, lograron puntuaciones de 3 y 4 puntos, con excepción de algunas valoraciones de 2 puntos, principalmente de un juez. Estos valores son consistentes con los ajustes realizados en los ítems más críticos (5, 11, 13, 15 y 16), ya que sus puntuaciones aumentaron en la segunda evaluación.

### Contextualización para el razonamiento inferencial informal

En este criterio se identificaron desafíos importantes en la primera evaluación. La Tabla 13 muestra que doce ítems recibieron al menos una valoración inferior a 3 por parte de los jueces expertos, indicando un problema en los contextos presentados para activar el razonamiento inferencial informal en los estudiantes de 7 a 10 años.

**Tabla 13**

*Puntuaciones obtenidas de ítems críticos y en revisión para el criterio de “Contextualización para el razonamiento inferencial informal” en la primera y segunda evaluación de jueces expertos*

Jueces	Ítem	Contextualización para el razonamiento inferencial informal	
		Primera evaluación	Segunda evaluación
M	1	2	2
M	2	2	2
J	3	NA	4
J	5	2	4
M	5	2	3
R	5	2	4
M	6	2	3
M	7	2	2
M	9	2	3
R	9	2	4
M	11	2	4
M	12	2	2
M	13	2	2
R	13	1	3
J	15	2	4
J	16	2	4

*Nota.* El rojo indica ítems con puntajes críticos y el rosado ítems que requieren revisión. El amarillo corresponde a aquellos puntajes que subieron. Las celdas sin color corresponden a puntuaciones que se mantuvieron sin cambios.

El análisis revela que los ítems 1, 2, 5, 6, 7, 9, 11, 12, 13, 14 y 16 presentaron valoraciones de 2 puntos, mientras que el ítem 13 obtuvo la calificación más baja con puntaje 1. El ítem 3 no recibió calificación por parte de uno de los jueces. Esto evidencia que la formulación inicial de los contextos planteados no resultaba suficientemente familiares, significativos o adecuados para el grupo etario.

En la segunda evaluación se evidenció una mejora en la contextualización. De los doce ítems que presentaban problemas, ocho lograron puntuaciones de 3 y 4 puntos en la evaluación de la mayoría de los jueces. Sin embargo, cinco ítems mantuvieron valoraciones de 2 puntos, únicamente por parte de un juez, mientras que el resto les asignaron puntajes superiores.

Los resultados de la segunda evaluación (Tabla 13) confirman cuantitativamente las mejoras descritas. A pesar de que algunos ítems mantuvieron valoraciones de 2 puntos por parte de un juez, la mayoría de los ítems que presentaban problemas, con los cambios hechos, lograron puntuaciones de 3 y 4 puntos. Este patrón de resultados es coherente con los ajustes realizados en los ítems 3, 5, 6, 9, 11, 13, 15 y 16, que en la segunda evaluación obtuvieron puntuaciones de 3 y 4 puntos en la mayoría de los jueces.

### **Relevancia de las alternativas**

En este criterio se observaron diferencias significativas entre la primera y la segunda evaluación de jueces expertos. En la primera revisión, varios ítems obtuvieron puntuaciones de 2 puntos por parte de los jueces, lo que evidenciaba problemas en la coherencia entre las alternativas de respuesta y los enunciados. Este resultado indicaba que las alternativas no cumplían adecuadamente su función discriminativa y que, en algunos casos, los distractores incluían información irrelevante o demasiado evidente para los estudiantes.

Como se muestra en la Tabla 14, los ítems 2, 4, 5, 6, 7, 9, 12, 13, 15 y 16 presentaron al menos una valoración de 2 puntos o 1 en este criterio, mientras que el ítem 3 y 15 no obtuvieron valoración y el resto no obtuvo puntuaciones críticas. Las observaciones de los jueces apuntaron a distractores que no se relacionaban con la situación planteada, alternativas que no representaban errores inferenciales posibles para estudiantes de 7 a 10 años y opciones que podían descartarse sin recurrir a la sub-habilidad requerida por el ítem. En conjunto, estos problemas afectaban la pertinencia de las

alternativas y limitaban la capacidad para evaluar de manera rigurosa el pensamiento crítico.

**Tabla 14**

*Puntuaciones obtenidas en ítems críticos y en revisión para el criterio de “Relevancia de las alternativas” en la primera y segunda evaluación de jueces expertos*

Jueces	Ítem	Relevancia de las alternativas	
		Primera evaluación	Segunda evaluación
R	2	2	2
J	3	NA	4
C	4	2	2
R	4	2	2
J	5	2	4
R	5	2	4
M	6	2	2
J	7	2	4
J	9	3	4
M	9	2	4
J	11	3	4
M	11	3	4
C	12	2	4
M	12	2	2
R	13	2	3
J	15	1	4
M	15	2	2
R	15	NA	4
J	16	2	4
M	16	2	3

*Nota.* El rojo indica ítems con puntajes críticos y el rosado ítems que requieren revisión. El amarillo corresponde a aquellos puntajes que subieron. Las celdas sin color corresponden a puntuaciones que se mantuvieron sin cambios.

Tras la reformulación de los ítems con dificultades, se realizaron ajustes focalizados en la redacción y diseño de las alternativas. Estos incluyeron la eliminación de distractores inconsistentes, la clarificación de opciones cuya formulación llevaba a interpretaciones erróneas.

Los resultados de la segunda evaluación muestran una mejora evidente en este criterio. La mayoría de los ítems que inicialmente recibieron puntuaciones de 2 puntos

alcanzaron valores más altos, quedando sólo los ítems 2, 4, 6, 12 y 15 con al menos una valoración de 2 puntos.

### Cobertura del pensamiento crítico

En este criterio se observaron puntuaciones bajas por parte de los jueces en relación con la cobertura del pensamiento crítico. En la primera evaluación (Tabla 15), los dieciséis ítems recibieron calificaciones entre 1 y 3 puntos por parte de algunos jueces. Además, se registró una gran cantidad de valores “NA” en 12 de los ítems, lo que muestra que varios jueces optaron por no evaluar este criterio. Esta situación sugiere que los ítems, en general, no cubrían adecuadamente el pensamiento crítico o que no estaba alineada la relación entre habilidad y sub-habilidad.

**Tabla 15**

*Puntuaciones obtenidas en ítems críticos y en revisión para el criterio de “Cobertura del pensamiento crítico” en la primera y segunda evaluación de jueces expertos*

Jueces	Ítem	Cobertura del pensamiento crítico	
		Primera evaluación	Segunda evaluación
J	1	2	2
J	2	NA	3
J	3	NA	4
R	3	2	2
J	4	NA	4
J	5	NA	3
M	6	2	2
R	6	2	2
J	7	NA	3
N	7	2	2
R	8	2	2
J	9	NA	4
M	9	2	3
J	10	NA	4
R	10	2	2
J	11	NA	4
M	11	3	4
R	11	2	3
M	12	2	2
R	12	2	2

J	13	NA	2
R	13	1	3
J	14	NA	4
M	14	2	2
J	15	NA	3
M	15	1	1
R	15	NA	4
J	16	NA	4
M	16	1	3
R	16	2	4

*Nota.* El rojo indica ítems con puntajes críticos y el rosado ítems que requieren revisión. El amarillo corresponde a aquellos puntajes que subieron. Las celdas sin color corresponden a puntuaciones que se mantuvieron sin cambios.

En la segunda evaluación la situación cambia significativamente. De los doce ítems que presentaban “NA” en la primera evaluación, once de ellos pasaron a obtener calificaciones de 3 y 4 puntos, mientras que solo uno recibió una calificación de 2 por parte de un juez. Tras las correcciones y aclaraciones realizadas, los jueces lograron identificar mejor la relación entre el ítem y el pensamiento crítico. Sin embargo, algunos ítems mantuvieron puntuaciones bajas asignadas por un juez.

Gracias a las modificaciones realizadas, los ítems fueron mejor comprendidos y evaluados por los jueces, lo que permitió una adecuada cobertura del pensamiento crítico dentro del instrumento. Esto mostró un mayor consenso entre los jueces y una mayor claridad entre las habilidades y sub-habilidades que cada ítem buscaba medir.

#### 4.2.2 Análisis y ajustes realizados por ítem

A continuación, se presenta un análisis detallado de aquellos ítems a los que se le realizaron cambios, específicamente en aquellos criterios donde presentaron puntuaciones de 1 y 2 puntos, a partir de las evaluaciones de los jueces expertos.

##### Ítem 1

**Pertinencia.** Sólo el juez J de los otros siete jueces otorgó una calificación menor a tres puntos (2), debido a una confusión respecto a la sub-habilidad que el ítem buscaba evaluar, y no a un problema de formulación. En los comentarios realizados durante la

evaluación de jueces expertos, el juez J señaló que el ítem es pertinente, aunque consideró “difícil precisar si cubre por completo la sub-habilidad”. Tras aclarar el concepto de la sub-habilidad y el propósito del ítem, el juez decidió aumentar su puntuación por un 4. Considerando que los otros seis jueces valoraron positivamente este criterio desde la primera evaluación, no se consideró necesario realizar modificaciones estructurales.

**Contextualización para el razonamiento inferencial informal.** En la evaluación de este criterio, el juez M solo pudo evaluar el ítem 1 en la primera ronda, otorgándole 2 puntos. Al no estar disponible para su segunda evaluación, esta puntuación se mantuvo sin cambios. Sin embargo, el resto de los jueces otorgaron puntuaciones de 3 o 4 puntos, por lo que no se introdujeron modificaciones en el ítem.

**Cobertura del pensamiento crítico.** El juez J asignó una calificación de 2 puntos, la cual se mantuvo en ambas evaluaciones. En su comentario, el juez J señaló que “la cobertura del pensamiento crítico no es posible definirla con claridad”, sugiriendo hacer ajustes y entregar más información. Esta observación apunta principalmente a la necesidad de explicitar con mayor detalle, en el instrumento de validación enviado a los jueces expertos, cómo se operacionaliza la habilidad y la sub-habilidad en el ítem. El resto de los jueces otorgó puntuaciones de 3 y 4 puntos, lo que indica que, para la mayoría de los evaluadores, el ítem sí representa de manera suficiente la sub-habilidad evaluada. En consecuencia, se mantuvo la estructura del ítem.

## **Ítem 2**

**Claridad.** El criterio recibió una evaluación de 2 puntos por parte del juez R, cuya puntuación se mantuvo sin cambios en la segunda evaluación. Esto se debió a que el juez R solo participó en la primera ronda de evaluación para el criterio de este ítem en particular. El juez R cuestionó la expresión “5 niños votaron la manzana como preferida”, indicando que no es explícito el procedimiento de recogida de datos: “¿votaron? ¿fueron las colaciones que llevaron?”. Esta observación apunta a una necesidad de entregar mayor detalle sobre el contexto de los datos. Dado que los otros seis jueces otorgaron

puntuaciones de 3 y 4 puntos, y no reportaron dificultades de comprensión, se consideró que el ítem mantiene un nivel de claridad suficiente, por lo que no se realizaron modificaciones.

**Contextualización para el razonamiento inferencial informal.** En este criterio, el juez M evaluó con un puntaje de 2 en la primera ronda, calificación que se mantuvo al no participar en la segunda evaluación para este criterio e ítem en particular. Dicho puntaje resultó poco representativo, ya que solo uno de los siete jueces otorgó dicha puntuación. El resto mantuvo puntuaciones de 3 o 4 puntos, lo que indica que, en general, los expertos consideraron que la situación planteada es correcta para niños de 7 a 10 años y ofrece información suficiente. La puntuación del juez M se interpretó como una opinión aislada y no como evidencia de un problema consistente de contextualización, por lo que no se realizaron ajustes al ítem desde este criterio.

**Relevancia de las alternativas.** El juez R entregó una puntuación de 2 puntos para este criterio, la cual se mantuvo sin cambios al no participar en la segunda evaluación de este criterio e ítem en particular. En sus comentarios, el juez R sostuvo que todas las alternativas eran posibles y sugirió reformular la alternativa d). Dado que los otros jueces evaluaron este criterio con 3 y 4 puntos, se concluyó que las alternativas resultan coherentes con el enunciado y con la sub-habilidad evaluada. Debido a esto, se tomó la decisión de no realizar cambios.

**Cobertura del pensamiento crítico.** En la evaluación del criterio, el juez J determinó, en una primera instancia, no asignar puntaje alguno por considerar que el criterio en cuestión no podía medir a través de este ítem. Posteriormente, y tras recibir una fundamentación sobre la forma en que el ítem sí evidenciaba dicha competencia, el juez reconsideró su postura y asignó una puntuación de 3 en la segunda ronda de evaluación.

### Ítem 3

**Claridad.** Inicialmente, el juez J no asignó puntuación a este criterio por considerar que el ítem era similar a otro y sugirió una revisión conjunta. De forma paralela

a la evaluación de los jueces, el análisis de las entrevistas cognitivas identificó que la principal dificultad radicaba en la repetición del término "Disney" en el enunciado y en una de las opciones de respuesta. Esta recurrencia actuó como un distractor potente, generando un sesgo de familiaridad que llevaba a los estudiantes a seleccionar la opción c) por un reconocimiento superficial del término y no mediante un razonamiento conceptual. El ajuste en este criterio se derivó, por tanto, de los hallazgos cognitivos. Tras la modificación, en la segunda ronda de evaluación el juez J asignó una puntuación de 4

**Pertinencia.** La mayoría de los jueces asignó puntuaciones de 3 y 4 puntos, lo que indica un acuerdo general respecto a su alineación con la habilidad y sub-habilidad que se pretende medir. Sin embargo, el juez J no asignó puntuación en este criterio, mientras que el juez M lo evaluó con 2 puntos en la primera evaluación, puntuación que se mantuvo al no participar en la segunda ronda. En los comentarios del juez M, señala que la tarea podría exigir un nivel de abstracción superior al esperado si los estudiantes no han trabajado previamente la noción de “cantidad suficiente de datos” o “muestra representativa”. Esta observación apunta a la demanda cognitiva del contenido, no a una coherencia entre el ítem y la sub-habilidad. Dado que el objetivo del ítem es precisamente que los estudiantes juzguen si la información disponible es suficiente para aceptar una conclusión, se consideró que la crítica se refiere al grado de dificultad, no a la pertinencia conceptual del ítem. Debido a que el resto de los jueces no identificó problemas en este criterio, no se realizaron modificaciones. En la segunda evaluación y considerando los cambios realizados por las entrevistas cognitivas, el juez J asignó un puntaje de 4.

**Contextualización para el razonamiento inferencial informal.** En la primera evaluación, el juez J no calificó este criterio, al considerar que el ítem era similar a otro. Tras la reformulación motivada por las entrevistas cognitivas, el ítem fue reevaluado con 4 puntos, lo que indica que la situación se percibe como adecuada y comprensible para niños de 7 a 10 años, por lo que no se realizaron ajustes adicionales.

**Relevancia de las alternativas.** El juez J no realizó una valoración inicial por la razón previamente indicada, mientras que el resto de los jueces otorgó puntuaciones entre

3 y 4 puntos. Dado que no se detectaron distractores irrelevantes ni ambigüedades, el ítem se mantuvo sin modificaciones. No obstante, durante la segunda ronda de evaluación (y considerando los cambios implementados a raíz de las entrevistas cognitivas) el juez J asignó una puntuación de 4.

**Cobertura del pensamiento crítico.** La mayoría de los jueces asignaron puntuaciones de 3 y 4 puntos en este criterio. No obstante, en la primera evaluación el juez J no lo valoró, mientras que el juez R otorgó 2 puntos, manteniendo esta calificación al no participar en la segunda ronda. En sus comentarios, el juez R señaló que el ítem introduce la distinción entre condición suficiente y necesaria, observando que "se trata de pensamiento formal" y que puede estar "muy cargado a la lógica". Esta apreciación alude al nivel de razonamiento requerido, más no a la ausencia de pensamiento crítico en el ítem. Considerando que el resto de los jueces no cuestionó la cobertura del constructo, se determinó que el ítem representa adecuadamente la sub-habilidad evaluada y, por lo tanto, no se realizaron ajustes. Finalmente, en la segunda evaluación, y tras incorporar los cambios derivados de las entrevistas cognitivas, el juez J asignó una puntuación de 4.

#### **Ítem 4**

**Pertinencia.** El juez M evaluó la pertinencia con 2 puntos en la primera evaluación, puntuación que se mantuvo al no participar en la segunda evaluación de este criterio e ítem en particular. Sin embargo, la mayoría (6 de 7) consideró que el ítem es adecuado. El comentario del juez señalaba que podría ser difícil por la idea de "representatividad", pero el ítem solo requiere que los niños reconozcan que una conclusión es poco confiable cuando se basa en muy pocas personas, algo comprensible para su edad. Como el ítem es claro, cotidiano y está alineado con la habilidad que se busca medir, no se consideraron necesarios ajustes.

**Relevancia de las alternativas.** Los jueces C y R otorgaron 2 puntos en la primera evaluación, puntuaciones que se mantuvieron al no participar en la segunda ronda. La mayoría (5 de 7) evaluó que las alternativas eran pertinentes y coherentes con el propósito

del ítem. Los comentarios críticos se enfocan en detalles de redacción (“voto”, “encuesta”) o en la posibilidad de mejorar alternativas específicas, pero no cuestionan la función esencial de las opciones, que es ofrecer razones correctas e incorrectas para evaluar la credibilidad de la conclusión.

Las alternativas permiten distinguir claramente entre justificaciones válidas y no válidas, por lo que cumplen adecuadamente con el criterio de relevancia. Por esta razón, y considerando el acuerdo mayoritario, no se consideró necesario realizar ajustes.

**Cobertura del pensamiento crítico.** En este criterio, uno de los jueces (J) no asignó puntaje en su primera revisión porque no lograba identificar cómo el ítem reflejaba la cobertura del constructo. Una vez aclarado el modo en que este criterio se operacionaliza y explicada la relación entre el ítem y la sub-habilidad evaluada, el juez ajustó su valoración a 4 puntos. Dado que la falta de puntaje se originó únicamente por una duda respecto del procedimiento, y que los otros seis jueces evaluaron este criterio de manera positiva y sin dificultades, no se consideró necesario realizar modificaciones al ítem.

A diferencia de la revisión de expertos, las entrevistas cognitivas sí mostraron un problema de comprensión: algunos estudiantes confundieron la palabra “jugo” con “juego”, lo que los llevó a interpretar el ítem de manera incorrecta y a razonar desde un contexto distinto al previsto. Esta confusión afectaba la claridad y la validez del ítem.

Por ello, aunque los jueces no identificaron necesidad de cambios, las entrevistas evidenciaron que era necesario ajustar el enunciado. Se reemplazó “jugo” por “bebida” para evitar ambigüedades y asegurar una interpretación consistente con la habilidad de pensamiento crítico que se busca evaluar.

## **Ítem 5**

El ítem 5 fue un caso crítico, sometido a numerosas modificaciones (Figura 13) en varios de sus criterios. Estos cambios respondieron a las bajas puntuaciones otorgadas

por los jueces expertos, lo que hace que su análisis sea más complejo en comparación con los demás ítems.


### Figura 13

Comparación entre el ítem 5 original y su modificación

**ÍTEM 5: ORIGINAL**

María preguntó en su curso qué sabores de helado les gusta más.

Título. Sabores de helados preferidos



¿Qué debe hacer María para resolver el problema?


- a) Convencer a sus compañeros de elegir el de fruta.
- b) Descubrir qué sabor tiene más votos.
- c) Decidir si comprar frutilla o chocolate.
- d) Averiguar si a los profesores les gusta el helado.

↓

**ÍTEM 5: MODIFICADO**

María quiere saber qué sabor de helado le gusta más a sus compañeros. Preguntó a algunos de ellos y obtuvo lo siguiente:

Título. Sabores de helados preferidos



¿Qué debe hacer María para encontrar la respuesta?

- a) Convencer a sus compañeros de elegir un mismo sabor.
- b) Descubrir qué sabor de helado eligieron más.
- c) Decidir si comprar helado de frutilla o chocolate.
- d) Averiguar si a los profesores les gusta el helado de chocolate.

*Nota.* En el ítem original, el resaltado en rojo señala los elementos que requerían ajuste. En el ítem modificado, el resaltado en verde resalta los cambios implementados.

**Claridad.** En la primera evaluación, los jueces J, M y R asignaron 2 puntos en este criterio. Las observaciones se centraron en dos aspectos: el uso de la expresión “resolver el problema” y la falta de precisión en relación entre el enunciado y el pictograma. El juez R señaló que “no diría problema, sino respuesta”, mientras que el juez M destacó que el enunciado “habla de un problema, pero no lo dice antes. De hecho, no hay problema...”. A partir de esos comentarios, se reemplazó la expresión “resolver el problema” por “encontrar la respuesta” en la pregunta y se incorporó la frase “y obtuvo lo siguiente” en

el enunciado para articular explícitamente el enunciado con el pictograma. Al realizar estos cambios y ser reevaluados por los tres jueces, estos aumentaron la puntuación a 4 puntos, lo que implica que se eliminó esa ambigüedad sobre la existencia de un problema.

**Pertinencia.** En pertinencia, sólo el juez J asignó una puntuación de 2 puntos, indicando que “no está del todo claro el por qué se les pregunta acerca de un problema que debe resolver María”. Tras la modificación de enunciado, referente a la sustitución de “resolver el problema” por “encontrar la respuesta” y la reformulación del enunciado, el juez J aumentó su valoración a 4 puntos. Los cambios de redacción permitieron alinear de forma más explícita el ítem con la sub-habilidad evaluada, por lo que no se consideró necesario realizar nuevas modificaciones en este criterio.

**Contextualización para el razonamiento inferencial informal.** En la primera evaluación, los jueces J, M y R calificaron este criterio con 2 puntos, señalando falta de contextualización y ambigüedades en la situación. El juez R cuestionó el uso de “votos” y la claridad del contexto, comentando que “no usaría voto, sino preferencia”. Tras ese comentario, se reemplazó la expresión “tienen más votos” por “eligieron más” en la alternativa b), y se incorporó la frase “y obtuvo lo siguiente” en el enunciado para vincular la acción de preguntar con el pictograma de sabores de helado. Estas modificaciones clarifican que se trata de una situación de preferencias dentro del curso y que la representación gráfica resume las respuestas de los compañeros, lo que refuerza la contextualización del ítem. Con los cambios realizados, en la segunda evaluación las puntuaciones de J y R aumentaron a 4 puntos y la de M a 3 puntos.

**Relevancia de las alternativas.** Dos jueces (J y R) evaluaron con 2 puntos este criterio. El juez R observó que “no es claro qué helado es de fruta”, que “no es claro el contexto de votación” y que la alternativa d), al referirse a los profesores, quedaba fuera de contexto. El juez J comentó en términos generales que las alternativas “no están tan coherentes”. Para solucionar eso, se cambió en la alternativa a) la palabra “fruta” por “un mismo sabor”, y se agregó la frase “helado de...” en la alternativa c). En la alternativa d) se agregó la frase “de chocolate” para aumentar la especificación en la alternativa. Tras los

cambios realizados, ambos jueces aumentaron su puntuación a 4 puntos en la segunda evaluación.

**Cobertura del pensamiento crítico.** En este criterio, un juez (J) no asignó puntuación en la primera evaluación. Sin embargo, luego de explicitar cómo el ítem operacionaliza la sub-habilidad evaluada, aumentó el puntaje a 3 puntos. Dado que la ausencia de puntuación se debió a una duda de interpretación del criterio, y considerando que tras las aclaraciones la valoración aumentó, no se realizaron modificaciones.

### **Ítem 6**

**Claridad.** Aunque dos jueces (M y P) otorgaron puntajes de 2 en claridad, la mayoría (5 de 7) consideró que el ítem es comprensible y utiliza un lenguaje adecuado para niños de 7 a 10 años. Las observaciones realizadas por los jueces que calificaron más bajo no se relacionan directamente con problemas de claridad del enunciado.

El juez M cuestionó la clasificación del ítem dentro del marco de la habilidad y sub-habilidad que se medían, señalando que podría corresponder a otra habilidad y sub-habilidad; sin embargo, esta observación apunta a la “pertinencia” o “cobertura del pensamiento crítico”, no a la comprensión del ítem por parte de los estudiantes.

Por su parte, el juez P indicó que la alternativa D tiene una sintaxis distinta a las demás, pero este comentario no afecta la claridad del enunciado principal ni la comprensión general de las opciones de respuesta. Dado que el ítem se entiende con facilidad, presenta un lenguaje accesible y no genera problemas de interpretación en su formulación, y considerando el consenso mayoritario positivo, no se consideraron necesarios ajustes en el criterio de claridad.

Sin embargo, luego de realizar esas aclaraciones, el juez M, cambió su puntaje a 3 en la segunda evaluación. El puntaje del juez P se mantuvo, ya que solo participó en la primera evaluación de este criterio e ítem específico.

**Pertinencia.** En la primera evaluación, el juez M asignó una puntuación de 2, decisión que mantuvo en la segunda ronda al considerar que la clasificación del ítem dentro de la habilidad y sub-habilidad asignadas era incorrecta, sugiriendo que correspondía a otra. Dado que esta fue la única calificación baja en el criterio (los otros seis jueces otorgaron puntuaciones altas), se decidió no realizar modificaciones al ítem en relación con este aspecto.

**Contextualización para el razonamiento inferencial informal.** En la primera evaluación, sólo el juez M de los otros 6 jueces, asignó 2 puntos a este criterio, argumentando que el ítem no correspondía a la habilidad y sub-habilidad señaladas, sino a otras. Se le aclaró que dicha observación estaba más relacionada con los criterios de "Pertinencia" o "Cobertura del pensamiento crítico por ítem", y que, en cambio, el ítem sí cumplía con la contextualización para el razonamiento inferencial informal. Esto se debe a que el estudiante debe evaluar la conclusión de Diego y justificar su validez con base en la evidencia recabada (8 de 10 personas). La tarea no se limita a reconocer un dato, sino que requiere inferir la solidez del procedimiento utilizado por Diego para llegar a su conclusión. El estudiante debe identificar que la conclusión es adecuada al apoyarse en la mayoría observada, lo que implica explicar y justificar un proceso de razonamiento a partir de pistas contextuales implícitas, como la proporción, el tamaño del grupo y el objetivo de Diego. Tras esta aclaración, en la segunda evaluación, el juez M ajustó su puntuación a 3.

**Relevancia de las alternativas.** Entre los seis jueces, sólo el juez M asignó una puntuación de 2 puntos a este criterio. Su evaluación se basó en la consideración de que el ítem no se correspondía con la habilidad y sub-habilidad asignadas, sino con otras, lo que en su opinión podría afectar la redacción de las alternativas y el enfoque para abordarlas. Por este motivo, mantuvo su calificación de 2 puntos en la segunda evaluación. Sin embargo, al tratarse de una valoración baja aislada frente a las puntuaciones más altas del resto de los jueces, se decidió no introducir modificaciones en el ítem con base en este criterio.

**Cobertura del pensamiento crítico.** En la primera evaluación, los jueces M y R otorgaron una puntuación de 2 puntos. El juez M justificó su calificación al considerar que el ítem no evaluaba la habilidad y sub-habilidad asignadas, sino otras diferentes. Por su parte, el juez R argumentó en sus comentarios que el ítem "justifica la conclusión, pero no el procedimiento o conjunto de pasos que lleva a ella". Estas puntuaciones se mantuvieron en la segunda evaluación: el juez M insistió en su postura inicial, mientras que el juez R no participó en la reevaluación de este criterio e ítem en particular. Dado que el resto de los jueces no presentó puntuaciones bajas ni cuestionó la correspondencia del ítem con la habilidad y sub-habilidad definidas, se determinó no realizar modificaciones en el ítem.

### **Ítem 7**

**Contextualización para el razonamiento inferencial informal.** Sólo el juez M asignó 2 puntos a este criterio, el cual se mantuvo, ya que no participó en la segunda evaluación de este criterio e ítem. Dado que fue la única valoración baja, no se realizaron ajustes. Sin embargo, las entrevistas cognitivas mostraron dificultades de comprensión vinculadas a elementos culturales y lingüísticos del ítem, lo que llevó a realizar modificaciones para mejorar su claridad y adecuación al contexto de los estudiantes.

**Relevancia de las alternativas.** Este criterio recibió sólo una valoración de 2 puntos por parte del juez J, quien señaló que la alternativa correcta debía considerar a las manzanas como foco principal, y no a los estudiantes, por lo que sugirió invertir la redacción para expresar que "las manzanas son las preferidas por los estudiantes del 3ºA". A partir de esa observación, se reformuló la alternativa correcta para situar a las manzanas como elemento central de la afirmación, y se ajustó la alternativa a), reemplazando "divertidas" por "saludables" para evitar asociaciones erróneas derivadas de los íconos sonrientes del pictograma. Tras ese cambio, el juez J otorgó una puntuación de 4 puntos, por lo que se consideró que las alternativas tenían un nivel adecuado de relevancia y coherencia con el enunciado.

**Cobertura del pensamiento crítico.** En este criterio, el juez J decidió no evaluarlo y N le otorgó 2 puntos en la primera evaluación; en la segunda, J le asignó 3 puntos, mientras que la puntuación de N se mantuvo, debido a que no participó en la reevaluación del criterio de este ítem en específico. Dado que el resto de los jueces lo evaluó con 3 y 4 puntos, y la asignación de puntaje por parte de J, se concluyó que el ítem cubre de forma suficiente el constructo.

### **Ítem 8**

**Cobertura del pensamiento crítico.** Sólo uno (R) de los siete jueces asignó una calificación de 2 puntos, la cual se mantuvo, debido a que no participó en la reevaluación del criterio de este ítem en específico. En su comentario, señaló dudas respecto de si el ítem evidenciaba efectivamente la cobertura del constructo, preguntándose si ésta aparece al considerar las frases “le dio comida” y “a buscar a Ana” presentes en las alternativas, concluyendo que le parecía que no “por la poca cantidad de datos”. Esta observación se interpreta como una crítica al carácter formal de la tarea y a la cantidad de información disponible, más que a la ausencia de procesos de pensamiento crítico en la resolución del ítem. Dado que todos los demás jueces asignaron puntuaciones de 3 y 4 puntos, se consideró que la valoración otorgada por R era aislada, por lo que no se realizaron modificaciones.

### **Ítem 9**

**Claridad.** En la primera evaluación, los jueces J y M identificaron problemas de claridad en el ítem. J asignó 3 puntos y señaló que la mención de "los 8 niños" complicaba la comprensión, mientras que M otorgó 2 puntos al considerar que contenía información innecesaria. Paralelamente, las entrevistas cognitivas revelaron dos dificultades adicionales: la expresión "helado de leche" era interpretada como "helado de vainilla" por los estudiantes, generando confusión conceptual, y el número "8" en una de las alternativas funcionaba como distractor numérico, llevando a elecciones erróneas basadas en elementos gráficos y no en el razonamiento.

Con base en estos hallazgos, se realizaron ajustes específicos. Para abordar las observaciones de los jueces, se agregó la frase "Luego de hacer una encuesta en la pizarra" para contextualizar el origen de los datos, y se eliminó la referencia redundante a "los 8 niños". Para resolver los problemas detectados en las entrevistas, se reemplazó "helado de leche" por "helado de chocolate", término más familiar para los estudiantes, y se suprimió el distractor numérico "8" de las alternativas.

En la segunda ronda, el juez J aumentó su puntuación a 4 puntos y el juez M a 3 puntos. La comparación entre ambas evaluaciones muestra un aumento en las puntuaciones de claridad del ítem.

**Pertinencia.** Sólo el juez M asignó 2 puntos, indicando que la referencia a la “fila más larga” podía confundir la longitud visual del pictograma con la cantidad de elementos, generando un sesgo perceptivo. Las modificaciones realizadas en claridad ayudan a centrar la tarea en la justificación de la conclusión de María a partir de información recopilada sistemáticamente, reforzando su alineación con la sub-habilidad evaluada. Debido a esto, M cambió su puntuación inicial y asignó 3 puntos, lo que indica una mejora de la pertinencia.

**Contextualización para el razonamiento inferencial informal.** En la primera evaluación, dos de los siete jueces (M y R) asignaron 2 puntos a este criterio. El juez R señaló que el contexto “se oscurece, no queda claro si los estudiantes estaban en filas para obtener helados o responder otra encuesta”. Por su parte, el juez M identificó un problema central en el enunciado: este confundía la longitud visual de la fila con la cantidad de elementos representados, lo que podía generar un sesgo perceptivo en los estudiantes. Según M, si se mantenía esta representación, los niños podrían interpretar "fila más larga" como un atributo gráfico y no como una relación numérica entre cantidades de datos, lo que interferiría con el propósito inferencial y analítico del ítem.

Los cambios realizados en el ítem, detallados previamente en el criterio de claridad, aclararon que el pictograma representaba los resultados de una encuesta aplicada

por María a su curso, estableciendo explícitamente una situación de recolección y análisis de datos. Esta reformulación permitió resolver tanto la ambigüedad contextual señalada por R como el riesgo de sesgo perceptivo identificado por M. Tras las modificaciones, en la segunda evaluación el juez R aumentó su puntuación a 4 puntos y el juez M a 3 puntos, lo que refleja una mejora significativa en la coherencia contextual y la validez inferencial del ítem.

**Relevancia de las alternativas.** En este criterio, el juez J asignó 3 puntos, comentando que las opciones eran ambiguas y que la alternativa b) requería revisión. Por otro lado, M asignó 2 puntos y cuestionó la falta de coherencia discursiva, señalando que no se explotaba a quién se dirigía María, ni con qué propósito hablaba, lo que debilitaba la función comunicativa de cada alternativa. En respuesta, se reformuló la pregunta para precisar el destinatario, quedando como: “¿Para qué María le dice eso al curso?”. Además, se ajustaron las alternativas para vincular explícitamente la intención comunicativa (explicar, mostrar, convencer) con la acción verbal de María dirigida al curso. Tras esos cambios, J y M aumentaron su puntuación a 4 puntos, indicando que las alternativas pasaron a ser coherentes con el enunciado.

**Cobertura del pensamiento crítico.** Inicialmente, el juez J no asignó puntuación en este criterio, y M asignó 2 puntos. Las dudas estaban vinculadas a la ambigüedad del contexto y a la falta de claridad en la intención comunicativa, lo que dificultaba percibir cómo el ítem operacionalizaba el constructo. Con los ajustes realizados en los criterios anteriores, la cobertura del constructo quedó mejor delimitada y se respondió a los comentarios realizados, por lo que ambos jueces aumentaron su puntuación, J a 4 y M a 3.

## Ítem 10

**Pertinencia.** El juez M otorgó un puntaje de 2 para este criterio, señalando que el ítem podría exigir un nivel de razonamiento meta inferencial superior al esperado para 7–10 años. Según ella, los niños a esta edad suelen hacer conexiones directas (como pensar

“si a un grupo le gusta, al otro también”) sin cuestionar si la evidencia es suficiente o válida para ambos grupos. Por lo tanto, considera que la tarea exige un razonamiento superior al esperado, especialmente si no han tenido práctica previa en evaluar que hace una idea creíble.

Sin embargo, aunque esta observación señala una posible dificultad en la aplicación, la mayoría de los jueces no reportó problemas de alineación conceptual o contextual en el ítem, no se consideró necesario realizar ajustes en esta dimensión. El puntaje inicial del juez M se mantuvo, ya que no participó en la segunda aplicación de la evaluación para el criterio e ítem.

**Cobertura del pensamiento crítico.** En la primera evaluación, el juez J no asignó puntuación a este criterio, mientras que el juez R otorgó 2 puntos. El juez J justificó su decisión argumentando que no identificaba una manifestación clara de pensamiento crítico en el ítem y consideraba complejo medirlo según la habilidad y sub-habilidad declaradas. Por su parte, el juez R señaló que el ítem priorizaba una "valoración de argumentos formales", enfocándose en el pensamiento lógico formal, y destacó que el pensamiento inferencial informal permite generalizaciones incluso ante posibles datos negativos, aspecto que, en su opinión, no estaba siendo cubierto por el ítem.

Tras aclarar que el ítem evalúa pensamiento crítico al requerir que el estudiante identifique una generalización incorrecta (juzgando la validez del argumento y detectando que no se puede concluir sobre un grupo sin evidencia directa) el juez J reevaluó su postura y aumentó su puntuación a 4 en la segunda ronda. El juez R no participó en esta segunda evaluación, por lo que su calificación se mantuvo en 2 puntos. Dado que solo dos de los siete jueces habían asignado inicialmente puntuaciones bajas, y uno de ellos rectificó su evaluación tras la aclaración, se determinó no realizar modificaciones en el ítem.

## Ítem 11

Debido a las bajas puntuaciones de los expertos, el ítem 11 (un caso crítico) fue objeto de numerosas modificaciones en sus criterios (Figura 14), por lo que su análisis resulta más complejo que el de los demás.

### Figura 14

*Comparación entre el ítem 11 original y su modificación*

**ÍTEM 11: ORIGINAL**

El profesor del 4ºB quiere comprar dos tipos de colaciones para una fiesta. Para decidir, preguntó a 12 de los 14 de su curso y anotó los votos:

Frutas	Votos
Plátano	3
Manzana	2
Papas Fritas	3
Galletas	1

Si el profesor quiere que la mayoría del curso esté contento, ¿cuáles dos colaciones deberían elegir?

- Las dos más sanas, aunque tengan pocos votos.
- Plátano y Papas Fritas, porque son las que más votos obtuvieron.
- Una fruta y una golosina, para tener variedad.
- Las que, a él, como profesor, le gustan más.

**ÍTEM 11: MODIFICADO.**

El profesor del 4ºB quiere comprar dos tipos de colaciones para una fiesta. Para decidir, preguntó a algunos estudiantes de su curso por su colación preferida y anotó los votos:

Colaciones	Votos
Plátano	10
Manzana	4
Papas Fritas	8
Galletas	2

Si el profesor quiere que la mayoría de estudiantes de su curso estén contentos, ¿cuáles dos colaciones debería comprar?

- Las dos más saludables, aunque juntas tengan menos votos.
- Plátano y Papas Fritas, porque también podría preferirlas el resto.
- Una fruta y una galleta, para tener variedad.
- No podemos saberlo, porque no preguntó a todos.

*Nota.* En el ítem original, el resaltado en rojo señala los elementos que requerían ajuste. En el ítem modificado, el resaltado en verde resalta los cambios implementados.

**Claridad.** En este criterio, el juez J asignó 3 puntos, señalando que la redacción de la pregunta no era del todo clara y que faltaba coherencia. En paralelo, M asignó 2 puntos, advirtiendo que el encabezado de la tabla no correspondía con los datos presentados, lo que generaba confusión. A partir de estas observaciones, se reemplazó el

encabezado “Frutas” por “Colaciones” y se ajustó la redacción de la pregunta para explicitar que “el profesor quiere que la mayoría de los estudiantes de su curso estén contentos al elegir las colaciones”. Tras la reformulación, ambos jueces aumentaron sus puntuaciones a 4, lo que indica que las modificaciones resolvieron los problemas detectados.

**Pertinencia.** Tres jueces asignaron puntuaciones de 2 puntos en este criterio. J comentó que cuesta distinguir el alcance de la sub-habilidad; M señaló que el conjunto de datos era muy reducido y con diferencias mínimas en las frecuencias presentadas en el gráfico y R indicó que no ve planes diferentes, y que el ítem no es tan pertinente. En respuesta, se modificó la tabla de datos, aumentando el número de estudiantes consultados y se ajustaron las frecuencias para establecer diferencias más claras entre las opciones. Como resultado de los ajustes realizados, J y R aumentaron su puntuación a 3 puntos y M a 4, lo que muestra una mejora en la pertinencia del ítem.

**Contextualización para el razonamiento inferencial informal.** Sólo el juez M asignó 2 puntos en este criterio, señalando que el conjunto de datos es muy reducido y que las diferencias mínimas entre frecuencias podían limitar el razonamiento inferencial informal. Las modificaciones realizadas en la tabla fortalecen el contexto de una decisión realista del profesor sobre qué colaciones comprar para una fiesta del curso. Esto facilita que los estudiantes comparen las cantidades y elaboren inferencias sobre qué combinación satisface mejor a la mayoría. En la segunda evaluación, M aumentó su puntuación a 4 puntos, lo que indica que los ajustes realizados permitieron un contexto para el razonamiento inferencial informal más adecuado.

**Relevancia de las alternativas.** En este criterio, dos (J y M) de los siete jueces asignaron 3 puntos, señalando la necesidad de mejorar la redacción de algunas alternativas, ya que no resultaban del todo claras. A partir de estas observaciones, se realizaron ajustes puntuales: en la alternativa a) se sustituyó “sanas” por “saludables”, y “pocos votos” por “menos votos”; en la alternativa b) se reformuló la justificación para indicar incertidumbre; en la alternativa c) se reemplazó “golosina” por “galleta” y la

alternativa d) se cambió completamente, pasando de “Las que, a él, como profesor, le gustan más” por “No podemos saberlo, porque no preguntó a todos”. Tras esas modificaciones, ambos jueces aumentaron su puntuación a 4 puntos, lo que indica que las alternativas mejoraron, siendo más claras, coherentes y adecuadas.

**Cobertura del pensamiento crítico.** En este criterio, el juez J decidió no evaluar, mientras que M y R calificaron con 3 y 2 puntos respectivamente. J sugirió que, dado que los estudiantes debían hacer conjeturas, podría incluirse explícitamente la idea de alimentación saludable, y R indicó que no veía con claridad la pertinencia del ítem, recomendando desarrollar una variedad de planes diferentes. Tras aclarar a los jueces cómo se operacionaliza la sub-habilidad en este ítem y luego de los ajustes realizados en la tabla y las alternativas, la percepción de cobertura mejoró. En la segunda evaluación, J asignó 4 puntos, M aumentó su puntuación a 4 puntos y R a 3 puntos.

## **Ítem 12**

**Claridad.** El juez M asignó 2 puntos en este criterio, cuestionando la coherencia entre el enunciado y los datos, ya que la formulación “por qué algunos compañeros llegan tarde a clases” no correspondía a la información representada. A partir de este comentario, se reformuló el enunciado original y se reemplazó por “con qué medios de transporte es más fácil llegar tarde a clases”. Con ello se deja de atribuir causalidad directa a los medios de transporte, centrándose en identificar un patrón en los datos de quienes llegaron tarde, coherente con lo que muestra el gráfico. A pesar de las modificaciones realizadas, se mantuvo el puntaje, ya que el juez M no participó en la segunda evaluación de este criterio e ítem específico.

**Pertinencia.** En este criterio, durante la primera evaluación, el juez M otorgó 2 puntos, manteniendo ese puntaje en la segunda evaluación, ya que el juez no participó de la reevaluación en este criterio e ítem. Señaló que el ítem está clasificado bajo la sub-habilidad “Descripción de resultados”, pero en la práctica demanda interpretar y valorar la conclusión de Martina más que describir los datos, lo que podría desplazar el foco hacia

habilidades analíticas o evaluativas. Frente a esto, se revisó la relación entre el ítem y la sub-habilidad evaluada, y considerando que fue la única valoración baja, se decidió por no realizar modificaciones.

**Contextualización para el razonamiento inferencial informal.** En la primera evaluación, el juez M asignó 2 puntos a este criterio, puntuación que se mantuvo al no participar en la reevaluación. Su principal observación fue que el ítem no evaluaba específicamente la "descripción de resultados", sino que requería interpretar una conclusión. Las modificaciones realizadas en el ítem, detalladas en los otros criterios, abordaron estas observaciones y resolvieron las inquietudes planteadas para este criterio en particular.

**Relevancia de las alternativas.** Los jueces C y M asignaron una puntuación inicial de 2 a este criterio. El juez C sugirió uniformizar la estructura de todas las alternativas. Considerando esta observación y los comentarios previos del juez M, se realizaron dos tipos de ajustes: primero, se modificó la pregunta del ítem reemplazando "mejor" por "de otra forma", quedando como "¿Cómo puedes explicar de otra forma la conclusión de Martina?"; segundo, se reformularon las alternativas para mejorar su coherencia contextual. Aunque no se implementó completamente la uniformidad estructural sugerida, en las alternativas a), b) y c) se eliminó la palabra "Que" del inicio, y se sustituyó la alternativa d) original por "Los que viajan a pie o en bicicleta, llegan temprano a clases".

Estos cambios lograron una mejor articulación entre el enunciado, la pregunta y las opciones de respuesta. Como resultado, el juez C elevó su puntuación a 4, valorando la mejora contextual. El juez M mantuvo su calificación de 2 al no participar en la reevaluación, aunque las modificaciones también abordaban sus observaciones anteriores.

**Cobertura del pensamiento crítico.** Los jueces M y R asignaron una puntuación de 2 en la primera evaluación, manteniendo estas calificaciones al no participar en la reevaluación. Sin embargo, las modificaciones aplicadas al ítem en los demás criterios

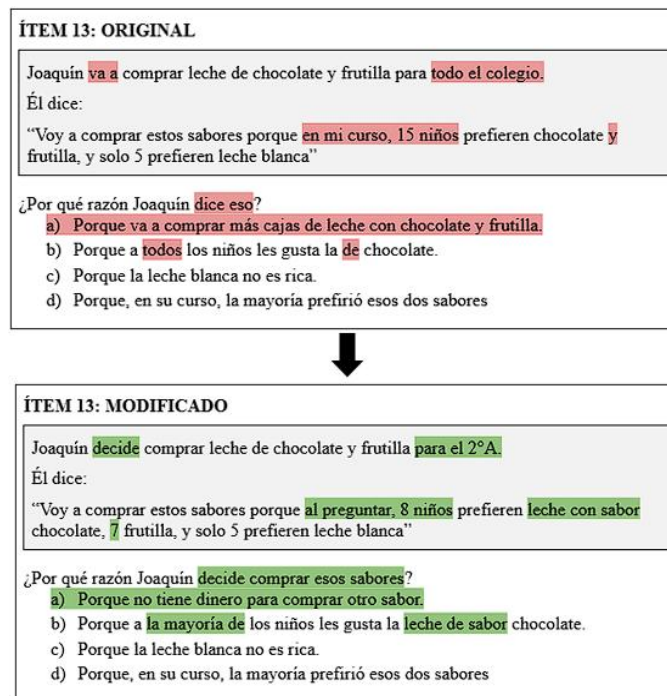
(como claridad, pertinencia, contextualización y relevancia de alternativas) abordaron de manera transversal las observaciones realizadas por ambos jueces sobre este criterio. Por lo tanto, sus puntuaciones iniciales ya no representan un aspecto pendiente de mejora en la versión final del ítem.

### Ítem 13

Las bajas puntuaciones de los jueces expertos convirtieron al ítem 13 en un caso crítico, lo que motivó revisiones sustanciales en varios de sus criterios (Figura 15) y, en consecuencia, un análisis notablemente más complejo.

### Figura 15

*Comparación entre el ítem 13 original y su modificación*



*Nota.* En el ítem original, el resaltado en rojo señala los elementos que requerían ajuste. En el ítem modificado, el resaltado en verde resalta los cambios implementados.

**Pertinencia.** Los jueces M y R asignaron inicialmente puntuaciones de 2 y 1, respectivamente, al ítem. El juez M señaló en sus comentarios que existía una

incongruencia en la situación planteada, lo que podía provocar confusiones o interpretaciones erróneas. Por su parte, el juez R indicó que el ítem parecía centrarse en medir razonamiento formal, pero que la redacción elevaba innecesariamente la dificultad para la edad objetivo.

Paralelamente, las entrevistas cognitivas revelaron que, aunque los estudiantes comprendían las palabras de forma aislada, la redacción redundante y extensa del enunciado dificultaba su comprensión global. Esto se evidenció en la incapacidad de los entrevistados para parafrasear el enunciado sin alterar su sentido.

Debido a las problemáticas identificadas tanto por los expertos como en las entrevistas, se realizaron modificaciones sustanciales en el enunciado. Se eliminaron frases como “va a”, “todo el colegio” y “en mi curso, 15 niños”, y se agregaron términos como “decide”, “para el 2° A”, “al preguntar”, “8 niños”, “leche con sabor” y “7”. El objetivo de estos ajustes fue mejorar la claridad, eliminar redundancias y ser más específico con los datos numéricos.

Tras las modificaciones realizadas en el criterio de “Pertinencia” y “Relevancia de las alternativas”, el juez R aumentó su puntuación a 3. La puntuación inicial del juez M se mantuvo, ya que no participó en la segunda evaluación. No obstante, esto no representa un problema grave, pues las principales inquietudes planteadas sobre el ítem pudieron subsanarse.

**Contextualización para el razonamiento inferencial informal.** En la primera evaluación, los jueces M y R otorgaron puntuaciones de 2 y 1, respectivamente. Ambos coincidieron en señalar que el enunciado era excesivamente extenso y presentaba una redacción inadecuada, lo cual incrementaba su dificultad. Además, criticaron que el ítem evaluaba razonamiento formal en lugar de razonamiento inferencial informal, y que las alternativas no se alineaban con este último constructo.

Estas observaciones fueron abordadas mediante la modificación de los criterios de “Pertinencia” y “Relevancia de las alternativas”. Tras revisar los ajustes realizados en el

ítem, el juez R elevó su calificación a 4 en la segunda evaluación. En el caso del juez M, su puntuación se mantuvo sin cambios, dado que no participó en esta fase de reevaluación.

**Relevancia de las alternativas.** El juez R asignó 2 puntos al criterio, señalando que las alternativas no eran precisas y carecían de coherencia con el enunciado. Aunque fue el único juez de los siete en identificar esta dificultad, se decidió proceder con ajustes en este aspecto. Esto se debió a que el juez M, al evaluar el criterio de “Contextualización para el razonamiento inferencial informal”, también destacó que las alternativas no eran adecuadas y sugirió fortalecerlas reformulándolas como razones: una basada en datos y otra en opinión.

Atendiendo estas observaciones, se realizaron las siguientes modificaciones en el ítem; en la pregunta: se reemplazó “dice eso” por “decide comprar esos sabores”. La alternativa a) fue completamente reformulada, pasando de “Porque va a comprar más cajas de leche con chocolate y frutilla” a “Porque no tiene dinero para comprar otro sabor”. En la alternativa b) se eliminaron las palabras “todos” y “de”, y se agregaron “la mayoría de” y “leche de sabor”.

Tras incorporar estas mejoras, derivadas de la revisión de los criterios de “Pertinencia” y “Relevancia de las alternativas”, el juez R elevó su puntuación a 3 en una reevaluación posterior.

**Cobertura del pensamiento crítico.** El juez R evaluó inicialmente este criterio con 1 punto, mientras que el juez J omitió su puntuación. Ambos coincidieron en que el ítem no cumplía adecuadamente con el criterio de cobertura del pensamiento crítico. Sin embargo, tras observar las modificaciones implementadas en el ítem (derivadas de los ajustes en los criterios de “Pertinencia” y “Relevancia de las alternativas”), se atendieron estas observaciones y se mejoró integralmente el ítem, logrando así evidenciar de manera más clara el pensamiento crítico. Como resultado, en la segunda evaluación ambos jueces actualizaron sus puntuaciones: el juez J asignó 2 puntos y el juez R otorgó 3 puntos.

## Ítem 14

**Claridad.** El juez M asignó 2 puntos en la primera evaluación, manteniendo esta calificación al no participar en la segunda ronda. Señaló que la pregunta sugería una certeza absoluta incompatible con la evaluación de justificaciones en contextos de incertidumbre, por lo que recomendó ajustar su redacción. Sin embargo, al tratarse de la única evaluación baja (los otros seis jueces valoraron positivamente el criterio), se determinó no modificar el ítem.

**Cobertura del pensamiento crítico.** Inicialmente, el juez J no asignó puntuación a este criterio, mientras que el juez M otorgó 2 puntos, calificación que mantuvo al no participar en la reevaluación. El juez M argumentó que el ítem no evaluaba adecuadamente la sub-habilidad de "presentar argumentos", ya que el estudiante solo debía seleccionar entre opciones predefinidas sin construir su propio razonamiento. Por su parte, el juez J consideró inicialmente difícil evidenciar el pensamiento crítico en el ítem.

Sin embargo, tras explicar que la tarea requiere evaluar una afirmación con evidencia limitada y distinguir justificaciones válidas de razones subjetivas, el juez J reconsideró su postura y asignó 4 puntos. Se aclaró que, si bien el formato de opción múltiple limita la construcción autónoma de argumentos, el ítem aborda esta habilidad al exigir identificar y seleccionar razones válidas entre diversas alternativas.

Dado que seis de los siete jueces asignaron puntuaciones altas al criterio (incluida la reevaluación del juez J), y solo el juez M mantuvo una calificación baja, se determinó no realizar modificaciones en el ítem.

## Ítem 15


El ítem 15 constituyó un caso crítico, dado que las bajas puntuaciones de los jueces expertos exigieron sucesivas modificaciones en sus criterios (Figura 16). Esta situación derivó en un proceso de análisis más complejo que para el resto de los ítems.

### Figura 16

*Comparación entre el ítem 15 original y su modificación*

**ÍTEM 15: ORIGINAL**

Martín **sacó una pelota de una caja y** dijo "todas las pelotas de la caja son azules"




**¿Qué debería haber pensado Martín antes de decir eso?**

- a) Que las pelotas de la caja **son** del mismo color.
- b) Que **el** color azul **es el más bonito**.
- c) **Solo sacó** una pelota, pero podría haber pelotas de otro color.
- d) **Como sacó** una pelota azul, las otras **son del mismo color**.

↓

**ÍTEM 15: MODIFICADO**

Martín **metió la mano en una caja con pelotas y, sin mirar, sacó una pelota azul.** **Entonces** dijo "todas las pelotas de la caja son azules"



**Si estuvieras en el lugar de Martín, ¿qué hubieras pensado al sacar la pelota azul?**

- a) Que las pelotas de la caja **podrían ser** del mismo color.
- b) Que **es la única pelota de** color azul **en la caja**.
- c) **Saqué** una pelota azul, pero **podrían** haber pelotas de otro color.
- d) **Saqué** una pelota azul, las otras **también tienen que ser azules**.

*Nota.* En el ítem original, el resaltado en rojo señala los elementos que requerían ajuste. En el ítem modificado, el resaltado en verde resalta los cambios implementados.

**Pertinencia.** Inicialmente, los jueces J y M asignaron 2 puntos a este criterio. El juez J señaló que la pregunta debía personalizarse hacia el estudiante para promover la

metacognición, mientras que el juez M indicó que el ítem no exigía una reflexión sobre el propio razonamiento.

Para abordar estas observaciones, se modificó el enunciado original: "Martín sacó una pelota de una caja y dijo 'todas las pelotas de la caja son azules'" por uno que explicitara el contexto de incertidumbre: "Martín metió la mano en una caja con pelotas y, sin mirar, sacó una pelota azul. Entonces dijo 'todas las pelotas de la caja son azules'". Asimismo, se reemplazó la pregunta original: "¿Qué debería haber pensado Martín antes de decir eso?" por una que situara al estudiante como protagonista: "Si estuvieras en el lugar de Martín, ¿qué hubieras pensado al sacar la pelota azul?".

Estos ajustes lograron personalizar la situación, fomentar la metacognición y enfatizar el razonamiento propio, tal como sugerían los jueces. Tras las modificaciones, el juez J elevó su puntuación a 4 en la segunda evaluación, mientras que el juez M mantuvo su calificación inicial al no participar en la reevaluación.

**Contextualización para el razonamiento inferencial informal.** Inicialmente, el juez J asignó 2 puntos a este criterio, señalando la necesidad de ajustar el ítem para alinearlos con la habilidad y sub-habilidad establecidas, así como para evidenciar procesos metacognitivos. Estas observaciones fueron abordadas mediante las modificaciones realizadas en los criterios de "Pertinencia" y "Relevancia de las alternativas", las cuales incorporaron un contexto de incertidumbre y situaron al estudiante como protagonista de la situación, favoreciendo la reflexión metacognitiva y la alineación del ítem con su respectiva habilidad y sub-habilidad. Tras revisar los ajustes aplicados al ítem, el juez J elevó su puntuación a 4 en la segunda evaluación.

**Relevancia de las alternativas.** En primera instancia, los jueces J y M asignaron puntuaciones de 1 y 2 puntos respectivamente, mientras que el juez R no emitió evaluación. el juez J reiteró la necesidad de alinear el ítem con la habilidad y sub-habilidad definidas, así como de evidenciar metacognición. El juez M señaló que las alternativas

evaluaban una falla inferencial más que un autoexamen cognitivo, y el juez R indicó que su redacción era inadecuada y podría inducir errores.

Para abordar estas observaciones, se reformularon todas las alternativas: en la opción a) se reemplazó "son" por "podrían ser"; en la b) se añadió "es la única pelota de... en la caja" y se eliminaron referencias subjetivas; en la c) y d) se incorporaron verbos en primera persona y expresiones que enfatizan el razonamiento bajo incertidumbre.

Tras los ajustes, los jueces J y R elevaron sus puntuaciones a 4, mientras que el juez M mantuvo su calificación inicial al no participar en la reevaluación.

**Cobertura del pensamiento crítico.** Los jueces J y R no asignaron puntuación a este criterio, mientras que el juez M otorgó 1 punto. Los jueces coincidieron en señalar que el ítem no evaluaba la habilidad y sub-habilidad definidas, no promovía la metacognición, su redacción podría inducir errores, y no situaba al estudiante como protagonista del razonamiento (como comentaban en los otros criterios). Estas observaciones fueron abordadas mediante las modificaciones realizadas en los criterios de "Pertinencia" y "Relevancia de las alternativas", las cuales resolvieron las problemáticas identificadas. Tras revisar los ajustes aplicados al ítem, los jueces J y R asignaron puntuaciones de 3 y 4 puntos respectivamente en la segunda evaluación. El juez M mantuvo su calificación inicial al no participar en la reevaluación.

### **Ítem 16**


La complejidad del análisis del ítem 16 se debe a su condición crítica, pues las bajas puntuaciones de los expertos obligaron a reformularlo reiteradamente (Figura 17), introduciendo cambios sustanciales en varios de sus criterios.

## Figura 17

### Comparación entre el ítem 16 original y su modificación

**ÍTEM 16: ORIGINAL**

Alejandra sacó una pelota roja de una caja y dijo "todas las pelotas de la caja **son rojas**"




Alejandra se dio cuenta de que fue muy apresurada ¿Qué **podría hacer para corregir lo que dijo?**

- a) Sacar varias pelotas y ver de qué color son.
- b) Cambiar su respuesta a "verde" sin sacar más pelotas.**
- c) Sacar una pelota más y si **es** roja, todas son rojas.
- d) Insistir en que **tiene la razón porque la primera pelota fue roja.**

↓

**ÍTEM 16: MODIFICADO**

Alejandra sacó una pelota roja de una caja y dijo "todas las pelotas de la caja **podrían ser** rojas"



Alejandra se dio cuenta de que fue muy apresurada ¿Qué **hubieras hecho tú en su lugar?**

- a) Sacar varias pelotas y ver de qué color son.
- b) Sin sacar más pelotas, decir que también podrían haber pelotas de otros colores.**
- c) Sacar una pelota más y si **sale** roja, todas **las demás** son rojas.
- d) Insistir en que **todas las pelotas de la caja son rojas.**

*Nota.* En el ítem original, el resaltado en rojo señala los elementos que requerían ajuste. En el ítem modificado, el resaltado en verde resalta los cambios implementados.

**Claridad.** Inicialmente, el juez J evaluó el criterio con 2 puntos, ya que, la versión inicial presentaba una afirmación absoluta que cerraba la posibilidad de cuestionamiento, lo que reducía la demanda cognitiva. Por esa razón se cambió la pregunta "¿Qué podría hacer para corregir lo que dijo?" por "¿Qué hubieras hecho tú en su lugar?". Esta modificación ayuda a mejorar la estructura lógica del enunciado, ya que la versión inicial

ubicaba la pregunta en un juicio externo sobre la acción de Alejandra y dejaba al estudiante como un observador pasivo, mientras que la pregunta reformulada obliga al estudiante a asumir el rol de quien cometió la conclusión apresurada y a trabajar la sub-habilidad evaluada: autocorrección.

Además de ese ajuste, se reemplazó “son rojas” por la expresión “podrían ser rojas” en el enunciado, dando un matiz de incertidumbre necesario para que el ítem active adecuadamente la sub-habilidad evaluada. Al dar espacio a la duda, se obliga al estudiante a reconocer que la conclusión inicial no está justificada. Estas modificaciones, permitieron que, en la segunda evaluación, el juez J modificara su puntaje a 4.

**Pertinencia.** Los jueces J y M, otorgaron una puntuación de 2 en la primera evaluación. El juez J mencionaba la necesidad de personalizar la pregunta hacia el estudiante que responde, enfocándose como un proceso personal más que en el juicio sobre el actuar de otros. En cambio, el juez M señalaba que la formulación del ítem y su contexto, no evaluaban como tal la habilidad y sub-habilidad que se pretendía.

A raíz de los comentarios, se incorporó la frase “hubieras hecho tú en su lugar” y se eliminó “podría hacer para corregir lo que dijo”. Esta modificación permitirá que el estudiante enfrente directamente el desafío de la sub-habilidad de autocorrección. Debido a las modificaciones en el ítem, en la segunda evaluación, ambos jueces aumentaron sus puntajes a 3 y 4.

**Contextualización para el razonamiento inferencial informal.** Inicialmente, el juez J asignó una puntuación de 2 a este criterio. Los ajustes realizados en los criterios de "Claridad" y "Pertinencia" (que incluyeron personalizar el escenario y acercarlo a la experiencia del estudiante) permitieron mejorar la contextualización de la situación de autocorrección planteada en el ítem. Tras presentar estas modificaciones al juez J y demostrar cómo el ítem ahora se adaptaba mejor al contexto y perspectiva del estudiante, el juez decidió elevar su puntuación a 4 en la segunda evaluación.

**Relevancia de las alternativas.** Los jueces J y M otorgaron una puntuación inicial de 2 en este criterio. Para mejorarlo, se rediseñaron las alternativas buscando personalizar la experiencia de los estudiantes. Se eliminó por completo la alternativa b) original: "Cambiar la respuesta a 'verde', sin sacar más pelotas" y se reemplazó por: "Sin sacar más pelotas, decir que también podrían haber pelotas de otros colores". En la alternativa c) se ajustó la redacción, incorporando términos como "sale" y "las demás" para hacerla más natural. Además, se modificó totalmente la alternativa d), que originalmente era "Insistir en que tiene razón porque la primera pelota fue roja", por una versión más coherente: "Insistir en que todas las pelotas de la caja son rojas".

Tras implementar estos cambios, ambos jueces reevaluaron el ítem: el juez J aumentó su puntuación a 4, y el juez M a 3, lo que refleja una mejora significativa en la relevancia y adecuación de las alternativas.

**Cobertura del pensamiento crítico.** En la primera evaluación, los jueces M y R asignaron puntuaciones bajas (1 y 2 puntos, respectivamente), mientras que el juez J omitió su calificación. Las modificaciones aplicadas al ítem en los criterios de claridad, pertinencia, razonamiento y relevancia (basadas en las observaciones de los jueces) permitieron abordar integralmente las inquietudes relacionadas con este criterio. Tras revisar la versión final del ítem, los tres jueces ajustaron sus evaluaciones: los jueces J y R asignaron 4 puntos, y el juez M, 3 puntos, reflejando una mejora sustancial en la cobertura del pensamiento crítico.

#### 4.2.3 Resultado de la validez de contenido según el coeficiente V de Aiken

Con el fin de estimar de manera cuantitativa la validez de contenido del instrumento, se calcularon los coeficientes de Aiken para cada ítem y para cada uno de los cinco criterios evaluados por los jueces expertos. En la Tabla 16 se presentan los valores obtenidos, junto con el promedio de V de Aiken por ítem.

**Tabla 16**  
*Resultados del coeficiente V de Aiken*

Ítem	Claridad	Pertinencia	Contextualización para el razonamiento inferencial informal	Relevancia de las alternativas	Cobertura del pensamiento crítico	Promedio de V de Aiken
1	0.762	0.952	0.810	0.810	0.762	0.819
2	0.810	0.905	0.857	0.857	0.952	0.876
3	0.810	0.905	0.905	0.952	0.810	0.876
4	0.857	0.857	0.905	0.714	0.952	0.857
5	0.952	0.952	0.905	1.000	0.905	0.943
6	0.762	0.810	0.905	0.857	0.762	0.819
7	0.952	0.905	0.857	0.857	0.810	0.876
8	0.952	0.905	0.952	0.857	0.857	0.905
9	0.857	0.952	0.905	0.857	0.905	0.895
10	0.810	0.905	0.905	0.905	0.857	0.876
11	0.810	0.857	0.810	0.952	0.952	0.876
12	0.810	0.905	0.905	0.762	0.810	0.838
13	0.857	0.810	0.857	0.857	0.810	0.838
14	0.762	0.905	0.905	0.810	0.905	0.857
15	0.857	0.905	0.905	0.857	0.762	0.857
16	0.905	0.952	0.952	0.857	0.952	0.924

*Nota.* Los valores resaltados en amarillo corresponden a puntajes de criterios específicos que se encuentran por debajo del puntaje de corte establecido ( $V \geq 0,80$ ), pero al ser puntajes aislados y que no afectan al promedio general del ítem no requieren revisión o eliminación.

De acuerdo con los criterios de interpretación definidos previamente en la Tabla 5 ( $V \geq 0,80 =$  ítem adecuado;  $0,70 \leq V < 0,80 =$  ítem en revisión;  $V < 0,70 =$  ítem a eliminar), todos los ítems del instrumento alcanzaron valores de V iguales o superiores a 0,819 en su promedio, lo que implica que ninguno quedó por debajo del puntaje de corte establecido para ser considerado adecuado.

Varios ítems destacaron por obtener puntajes altos, con valores que superaron el esperado o incluso alcanzaron 0.90 y más. Estos resultados demuestran que hubo un consenso sólido entre los jueces y una calificación muy positiva en los criterios de claridad, pertinencia, contextualización para el razonamiento inferencial, relevancia de las alternativas y cobertura del pensamiento crítico.

A pesar de que los promedios generales de la V de Aiken fueron altos, algunos ítems presentaron valores más bajos en algunos criterios. En la Tabla 16 se identifican puntajes de 0.762 en algunos criterios de los ítems 1, 4, 6, 12, 14 y 15. Estos valores no se encuentran dentro de la categoría de “adecuado” según los rangos establecidos, pero corresponden a criterios aislados y no al desempeño global del ítem. Por lo tanto, lo que realmente orienta la interpretación es el promedio final de la V de Aiken por ítem, los cuales todos superan el puntaje de corte, lo que significa que los ítems cumplen con el nivel de calidad esperado.

En conjunto, todos los ítems superaron el punto de corte definido para V de Aiken, y los puntajes críticos identificados en la primera evaluación aumentaron tras la reformulación de los ítems. En la segunda evaluación, los valores promedio por ítem se ubicaron entre 0,819 y 0,952.

## V. DISCUSIÓN

### 5.1 Validez del instrumento en el contexto de Educación Primaria

Los resultados que se obtuvieron en el presente estudio muestran coherencia con los antecedentes teóricos, ya que abordan los principales vacíos identificados en la literatura sobre pensamiento crítico en educación primaria: la casi inexistencia de instrumentos válidos y específicos para el primer ciclo básico, y el predominio de pruebas diseñadas para estudiantes mayores y con definiciones generales del constructo.

En primer lugar, los altos valores obtenidos en la V de Aiken demuestran que el instrumento posee una adecuada validez de contenido, lo que significa que sus ítems contribuyen de manera pertinente a la medición de las habilidades y sub-habilidades del pensamiento crítico en niños y niñas de 7 a 10 años. Esto atiende al vacío señalado por Ossa-Cornejo et al. (2017), quienes indicaban que la mayoría de los instrumentos existentes fueron diseñados para estudiantes universitarios o adultos, y solo algunos pocos para estudiantes de secundaria. De manera complementaria, Lai (2011) advierte que muchos instrumentos utilizados para evaluar el pensamiento crítico no fueron diseñados para aplicarse antes de cuarto grado, generando un vacío metodológico en la educación primaria y llevando a los profesores a usar herramientas que no corresponden. Frente a esto, los altos puntajes obtenidos por nuestro instrumento muestran que sí es posible construir una herramienta pertinente y ajustada al nivel cognitivo de los niños de educación básica.

Otros de los vacíos identificados en la literatura es la falta de instrumentos adecuados para medir el pensamiento crítico, especialmente en áreas como matemáticas, representa un desafío importante y que la ambigüedad en los instrumentos provoca que sea imposible distinguir si los profesores están midiendo el éxito en las habilidades de pensamiento crítico o solo las habilidades cognitivas normales generales (Evedi et al, 2023). En este estudio, los altos puntajes obtenidos en la V de Aiken en criterios como claridad, pertinencia, contextualización para el razonamiento inferencial informal,

relevancia de las alternativas y cobertura del pensamiento crítico indican que nuestro instrumento logra atender esta problemática, ya que permite evaluar específicamente las habilidades y sub-habilidades definidas por el informe Delphi (Facione, 1990), sin confundirse con procesos cognitivos generales.

Una dificultad adicional mencionada en la literatura tiene relación con la poca claridad en las evaluaciones de pensamiento crítico. Abrami et al. (2008) señalan que, cuando los objetivos o las instrucciones no están claramente definidas o las evaluaciones no son claras, resulta difícil conocer realmente el impacto de la instrucción. Pero, los resultados de nuestro instrumento, al mostrar indicadores positivos de validez y comprensión en la V de Aiken y en las entrevistas cognitivas, aportan a superar este problema, ya que ofrecen una herramienta que podría permitir a los docentes monitorear con mayor precisión el desarrollo del pensamiento crítico en estudiantes de primer ciclo básico en matemáticas.

Asimismo, la literatura también ha planteado que el pensamiento crítico suele asociarse a estudiantes de mayor edad. Massa (2014) explica que esta percepción se sostiene porque históricamente se ha vinculado a jóvenes y adultos. Sin embargo, las entrevistas cognitivas realizadas en nuestro estudio mostraron que los niños y niñas de 7 a 10 años pueden razonar críticamente frente a tareas contextualizadas y acordes a su nivel cognitivo. Esto coincide con lo planteado por Lin et al. (2017), quienes afirman que la educación primaria constituye un momento clave para fortalecer estas habilidades. En este sentido, nuestro estudio no sólo confirma esta postura, sino que también ofrece evidencia concreta de que el pensamiento crítico puede desarrollarse desde edades tempranas.

Los resultados obtenidos permiten comprender de manera más clara cómo el instrumento diseñado responde a los vacíos y problemáticas identificadas en la literatura. La evidencia obtenida en la validación de contenido muestra que es posible evaluar el pensamiento crítico en niños y niñas de educación primaria mediante tareas contextualizadas de acuerdo a su edad. Por lo tanto, los hallazgos obtenidos en este estudio

confirman que el instrumento responde a los vacíos detectados y que, además, constituye una propuesta concreta para avanzar en la evaluación del pensamiento crítico en estudiantes de primer ciclo básico, aportando una herramienta que antes no existía para este nivel escolar.

En síntesis, los resultados obtenidos no sólo aportan evidencia de validez de contenido del instrumento, sino que también abren posibilidades para su uso en el sistema escolar, especialmente como herramienta diagnóstica para docentes del primer ciclo de educación básica.

## **VI. CONCLUSIONES, LIMITACIONES Y PROYECCIONES DEL ESTUDIO**

### **6.1 Conclusiones sobre los objetivos y las preguntas de investigación**

El propósito de este estudio fue diseñar y realizar la validación de contenido de un instrumento cuantitativo para medir el pensamiento crítico en matemáticas para estudiantes de 7 a 10 años, en contextos de razonamiento inferencial informal. Esta investigación se organizó en torno a un objetivo general y tres objetivos específicos, guiados por cuatro preguntas de investigación.

A continuación, se sintetizan las conclusiones en relación con cada objetivo específico, qué se logró y cómo esos resultados permiten responder a las preguntas de investigación.

#### **6.1.1 Conclusiones sobre el objetivo general**

El objetivo general de esta investigación fue “desarrollar y realizar la validación de contenido de un instrumento de evaluación cuantitativo orientado a medir el pensamiento crítico en el área de matemáticas en estudiantes de entre 7 y 10 años de edad”. Para abordar este objetivo, se construyó un test de 16 ítems de opción múltiple, alineado con el modelo de pensamiento crítico de Facione, organizado en seis habilidades y sus sub-habilidades, contextualizado en tareas de razonamiento inferencial informal acordes al currículum nacional y al nivel cognitivo de estudiantes de 7 a 10 años.

Las entrevistas cognitivas mostraron que la mayoría de los ítems alcanzaba altos niveles de comprensión semántica y lingüística, comprensión de la tarea, comprensión conceptual, y de engagement y pertinencia percibida. Las dificultades detectadas se concentraron en un conjunto acotado de ítems, asociados principalmente a vocabulario ambiguo, contextos poco claros o distractores mal calibrados. Estas debilidades fueron corregidas mediante modificaciones específicas.

Paralelamente, el juicio de expertos y el cálculo del coeficiente V de Aiken evidenciaron niveles de acuerdo iguales o superiores al punto de corte definido en todos los ítems, lo que respalda la adecuación del instrumento para el rango etario considerado.

En síntesis, el objetivo general se cumple: se creó un instrumento teóricamente robusto y con evidencia consistente de validez de contenido para estudiantes de 7 a 10 años. Esto permitió responder de manera integrada a las preguntas de investigación: organiza los componentes conceptuales del pensamiento crítico en seis habilidades y sub-habilidades correspondientes; explicita los criterios para la construcción de instrumentos válidos y pertinentes en el contexto escolar chileno; y muestra que el juicio de expertos y las entrevistas cognitivas constituyen procedimientos complementarios y efectivos para establecer la validez de contenido de instrumentos que evalúen pensamiento crítico matemático en educación primaria.

#### 6.1.2 Conclusiones sobre el objetivo específico 1

El objetivo específico 1 de este estudio fue “identificar los instrumentos existentes para la medición del pensamiento crítico en el contexto educativo, particularmente en la educación primaria”.

La revisión de literatura mostró tres puntos críticos: (1) la mayoría de los instrumentos está dirigido a educación secundaria o superior; (2) en muchos casos, el constructo del pensamiento crítico se define de forma amplia o ambigua y, (3) en matemáticas, el pensamiento crítico tiende a evaluarse en pruebas de rendimiento general, sin una operacionalización explícita de habilidades de este constructo.

Como conclusión, el OE1 se cumple y permite responder de forma directa a la segunda pregunta de investigación: “¿Cuáles son los instrumentos o metodologías empleadas para evaluar el pensamiento crítico en matemáticas en estudiantes de primer ciclo, y qué características y limitaciones se identifican en ellos?”. La revisión realizada evidencia un vacío claro: no existen instrumentos específicos, teóricamente bien definidos y adaptados a niños de 7 a 10 años para evaluar el pensamiento crítico en contextos de

razonamiento inferencial informal, lo que justifica la necesidad del instrumento construido en este estudio.

Este análisis aporta una síntesis actualizada y crítica del estado de la evaluación del pensamiento crítico en educación primaria y respalda la necesidad de avanzar hacia instrumentos más específicos y conceptualmente definidos para el área de matemáticas.

### 6.1.3 Conclusiones sobre el objetivo específico 2

El objetivo específico 2 de este estudio fue “construir un instrumento cuantitativo para evaluar el pensamiento crítico matemático en estudiantes chilenos de 7 a 10 años”

Este objetivo se cumple. A partir de la operacionalización del modelo de Facione, se diseñó un instrumento de 16 ítems de opción múltiple, donde cada sub-habilidad de pensamiento crítico se vincula a un ítem específico. Los enunciados se situaron en contextos de razonamiento inferencial informal coherentes con el currículum nacional. Se realizaron ajustes de vocabulario, longitud de enunciados y apoyos visuales acordes al nivel de comprensión de estudiantes de 7 a 10 años. A la vez, se definieron distractores que representaran errores de razonamiento posibles, de modo que la tarea exigiera efectivamente la sub-habilidad que se estaba evaluando.

La construcción del instrumento aporta evidencia a la primera pregunta “¿Qué aspectos conceptuales componen el pensamiento crítico?”, al mostrar que los aspectos conceptuales del pensamiento crítico pueden operacionalizar en tareas específicas y observables en matemáticas. Además, responde de manera directa a la tercera pregunta “¿Qué criterios deben considerarse para diseñar un instrumento que evalúe de manera válida y pertinente el pensamiento crítico en matemáticas en el contexto escolar chileno?” al establecer que este debe sustentarse en un modelo teórico robusto, integrar el razonamiento inferencial informal, alinearse al currículum nacional y al nivel de desarrollo cognitivo de los estudiantes, contando con ítems que representen de forma clara cada sub-habilidad de pensamiento crítico

#### 6.1.4 Conclusiones sobre el objetivo específico 3

El objetivo específico 3 de este estudio fue “evaluar la validez del contenido del instrumento diseñado mediante el juicio de expertos y el análisis cualitativo de entrevistas realizadas a estudiantes”.

Este objetivo se cumple. Las entrevistas cognitivas mostraron que la mayoría de los ítems presentaban niveles altos en los cuatro criterios, lo que indicó que los estudiantes entendieron el vocabulario, las instrucciones y el tipo de razonamiento requerido, y se mantuvieron involucrados con las situaciones planteadas. También se evidenció que un conjunto reducido de ítems presentó dificultades, asociados a vocabulario ambiguo o poco familiar, contextos que generan interpretaciones erróneas y distractores que no representan correctamente los errores de razonamiento esperados. A partir de esas evidencias, se realizaron ajustes específicos para eliminar esas dificultades.

A su vez, los jueces expertos evaluaron cada ítem en cinco criterios: claridad, pertinencia, contextualización en razonamiento inferencial informal, relevancia de las alternativas y cobertura del pensamiento crítico. Este proceso se llevó a cabo mediante dos evaluaciones. En la primera, se identificaron ítems con baja puntuación en uno o más criterios, donde los comentarios de los expertos permitieron detectar problemas que debían corregirse. En base a dichas observaciones, se realizaron los ajustes necesarios para luego llevar a cabo la segunda evaluación, donde aquellos jueces que habían evaluado con puntajes bajos volvieron a evaluar los ítems reformulados.

Una vez terminadas las evaluaciones, se calculó la V de Aiken para cada ítem. Todos alcanzaron niveles iguales o superiores al punto de corte definido, lo que indica un alto nivel de acuerdo entre los jueces respecto a la calidad del contenido.

La combinación de las evaluaciones realizadas a través de las entrevistas cognitivas y los jueces expertos permiten concluir que el OE3 se cumple, pues el instrumento presenta evidencia consistente de validez de contenido. Además, responde a la cuarta pregunta de investigación “¿Cómo contribuyen el juicio de expertos y las

entrevistas cognitivas con estudiantes a establecer la validez de contenido del instrumento diseñado para evaluar el pensamiento crítico matemático en estudiantes de 7 a 10 años?” al evidenciar la pertinencia de combinar métodos cuantitativos y cualitativos en la validación de instrumentos para evaluar el pensamiento crítico en educación primaria.

En conjunto, estos resultados aportan un procedimiento replicable de validación de contenido que combina métodos cuantitativos y cualitativos y que puede ser utilizado como referencia en el diseño de nuevos instrumentos de evaluación del pensamiento crítico en niveles de educación primaria.

## **6.2 Limitaciones y proyecciones del estudio**

### 6.2.1 Limitaciones del estudio

El estudio presenta limitaciones que deben ser consideradas al momento de analizar los resultados.

Una de ellas, es el tamaño de la muestra. Las entrevistas cognitivas se aplicaron a un grupo de diez de niños y niñas, lo que permitió observar a gran detalle cómo comprendían los ítems. Sin embargo, este número limita la generalización de los resultados, ya que las formas de comprender o razonar pueden variar entre niños del mismo rango etario. Por esta razón, no es posible afirmar que las dificultades o soluciones identificadas sean representativas para los demás estudiantes.

Además, las entrevistas cognitivas fueron realizadas únicamente a estudiantes de la Región del Biobío y de La Araucanía, específicamente en las ciudades de Los Ángeles, Mulchén y Angol. Esta focalización geográfica restringe la diversidad de contextos escolares considerados, por lo que la comprensión de los ítems, los ajustes realizados y las conclusiones obtenidas podrían no ser representativas para estudiantes de otras regiones del país.

Por último, el instrumento se compone sólo de ítems de selección múltiple con una única respuesta correcta. Si bien este formato facilita la aplicación masiva y la corrección

estandarizada, limita la observación directa de procesos centrales del pensamiento crítico. En consecuencia, el test muestra principalmente el resultado de la decisión del estudiante a través de la alternativa elegida, pero no permite acceder en detalle a la calidad de sus explicaciones ni a las estrategias de razonamiento que utiliza. Por ello, en futuros estudios, sería deseable complementar este instrumento con tareas de desarrollo o respuestas abiertas que permitan evaluar de manera más completa ciertas habilidades y sub-habilidades del pensamiento crítico.

En consecuencia, los resultados y conclusiones de este estudio deben interpretarse con cautela y entenderse como un primer paso en el proceso de validación de un instrumento que requiere nuevas aplicaciones y análisis en muestras más amplias y diversas.

#### 6.2.2 Proyecciones del estudio

En términos de difusión, se proyecta socializar el instrumento y los resultados de este estudio con equipos directivos y docentes de establecimientos municipales y particulares subvencionados mediante jornadas de trabajo pedagógico y talleres de formación. Asimismo, se considera poner a disposición de pago el test, su manual de aplicación y orientaciones de interpretación a través de repositorios institucionales y plataformas digitales.

Estas proyecciones permitirían fortalecer la validez y confiabilidad del instrumento y ampliar sus posibilidades de uso en contextos educativos reales.

En primer lugar, el instrumento puede convertirse en una herramienta diagnóstica para docentes de primer ciclo de educación básica, permitiendo identificar fortalezas y debilidades en habilidades del pensamiento crítico en matemáticas, y, con ello, orientar la toma de decisiones pedagógicas.

En segundo lugar, futuros estudios pueden explorar la adaptación del instrumento a otros niveles escolares (p. ej., primero básico o segundo ciclo de educación básica) y

otros contextos curriculares (p ej., Ciencias Naturales), manteniendo la coherencia del modelo teórico utilizado en este estudio.

Finalmente, en el sentido psicométrico, hasta ahora el estudio se ha centrado exclusivamente en la validez de contenido; no se ha evaluado aún la validez de constructo ni la validez de criterio, por lo que el instrumento sólo cuenta con una evidencia inicial de validez. Se proyecta avanzar hacia un análisis psicométrico más robusto mediante una aplicación piloto con aproximadamente 100 estudiantes de escuelas municipales y particulares subvencionadas, que permita evaluar el comportamiento estadístico de los ítems. Posteriormente, se espera realizar una aplicación final con 400 estudiantes, con el fin de obtener evidencias más sólidas de validez y confiabilidad.

## REFERENCIAS

- Abrami, P. C., Bernard, R. M., Borokhovski, E., Wade, A., Surkes, M.A., Tamim, R., y Zhang, D. (2008). Instructional interventions affecting critical thinking skills and dispositions: A stage 1 meta-analysis. *Review of Educational Research*, 78(4), 1102–1134. <https://doi.org/10.3102/0034654308326084>
- Abrami, P. C., Bernard, R. M., Borokhovski, E., Waddington, D. I., Wade, C. A., y Persson, T. (2015). Strategies for teaching students to think critically: A meta-analysis. *Review of educational research*, 85(2), 275-314. <https://doi.org/10.3102/0034654314551063>
- Aiken, L. R. (1980). Content validity and reliability of single items or questionnaires. *Educational and Psychological Measurement*, 40, 955–959. <https://doi.org/10.1177/001316448004000419>
- Aracena, M., Gutiérrez, J., López, M., Valenzuela, C., Cuadra-Martínez, D., Oyanadel Véliz, C., Castro, P. J., y González Palta, I. (2024). Adaptación y validación de la Student Teacher Professional Identity Scale en estudiantes de pedagogía chilenos. *Interdisciplinaria*. <https://www.redalyc.org/articulo.oa?id=18076225009>
- Bellei, C. (2013). El estudio de la segregación socioeconómica y académica de la educación chilena. *Estudios pedagógicos*, 39(1), 325-345. <https://dx.doi.org/10.4067/S0718-07052013000100019>
- Ben-Zvi, D. (2006). Scaffolding students informal inference and argumentation. [https://www.ime.usp.br/~abe/ICOTS7/Proceedings/PDFs/InvitedPapers/2D1\\_BE\\_NZ.pdf](https://www.ime.usp.br/~abe/ICOTS7/Proceedings/PDFs/InvitedPapers/2D1_BE_NZ.pdf)
- Bernard, R. M., Zhang, D., Abrami, P. C., Sicol, F., Borokhovski, E., y Surkes, M. S. (2008). Exploring the structure of the Watson–Glaser Critical Thinking Appraisal: One scale or many subscales? *Thinking Skills and Creativity*, 3(1), 15-22. <https://doi.org/10.1016/j.tsc.2007.11.001>
- Caicedo-Cavagnis, E., Zalazar, M. F. (2018). Entrevistas cognitivas: Revisión, directrices de uso y aplicación en investigaciones psicológicas. *Avaliação Psicológica*, 17(3), 362-370. <https://doi.org/10.15689/ap.2018.1703.14883.09>
- Cañongo, A., Acle, G. y Reyes, N. G. (2020). Habilidades de nivel inferencial y crítico en alumnos de primaria. *Revista Electrónica de Investigación Educativa*, 22(12), 1-12. <https://doi.org/10.24320/redie.2020.22.e12.2364>
- Casas, M., Bosch, D., y González, N. (2005). Las competencias comunicativas en la formación democrática de los jóvenes: describir, explicar, justificar, interpretar y argumentar. *Enseñanza de las Ciencias Sociales*, (4), 39-52. <https://www.redalyc.org/pdf/3241/324127616006.pdf>

- Centro de Investigaciones Innocenti. (2013). *Carta internacional para la investigación ética de la niñez*. UNICEF.
- Choque, K. (2019). Desarrollo del pensamiento crítico en niños de educación primaria. *Asociación Centro de Investigaciones y Escuela de Pensamiento Crítico Aymara*, 1(1), 47-64. <https://doi.org/10.56736/2019/7>
- De Bie, H., Wilhelm, P. y van der Meij, H. (2015). La evaluación del pensamiento crítico de Halpern: Hacia una evaluación holandesa del pensamiento crítico. *Habilidades de pensamiento y creatividad*, 17, 33-44. <https://doi.org/10.1016/j.tsc.2015.04.001>
- Deng, M., Ma, J., Lv, X. & Ren, X. (2023). Academic performance and parenting styles differentially predict critical thinking skills and dispositions among primary students: Cross-sectional and cross-lagged evidence. *Thinking Skills and Creativity*, 50. <https://doi.org/10.1016/j.tsc.2023.101384>
- Dewey, J. (1910). *How we think*. D. C. Heath & CO., Publishers. <https://dn790008.ca.archive.org/0/items/howwethink00deweiala/howwethink00deweiala.pdf>
- Elder, L., y Paul, R. (2019). The thinker's guide to analytic thinking: How to take thinking apart and what to look for when you do. *Rowman & Littlefield*. <https://books.google.es/books?hl=es&lr=&id=NDqYDwAAQBAJ&oi=fnd&pg=PA2&dq=Paul+y+Elder.+The+Thinker%27s+Guide+to+Analytic+Thinking&ots=CE557Eqemd&sig=RZVOfqobX7JE5D-sctToJDYu0vM#v=onepage&q=Paul%20y%20Elder.%20The%20Thinker's%20Guide%20to%20Analytic%20Thinking&f=false>
- Ennis, R. H. (1962). A Concept of Critical Thinking. *Harvard Educational Review*. Volume 32 (1). [https://criticalthinking.net/wp-content/uploads/2024/04/Ennis\\_HER\\_v32n1\\_pp81-111.pdf](https://criticalthinking.net/wp-content/uploads/2024/04/Ennis_HER_v32n1_pp81-111.pdf)
- Ennis, R. H. (2018). Critical thinking across the curriculum: A vision. *Topoi*, 37, 165-184. <https://doi.org/10.1007/s11245-016-9401-4>
- Ennis, R. H., y Weir, E. (1985). *The Ennis-Weir Critical Thinking Essay Test: An instrument for teaching and testing*. Midwest Publications. <https://evolkov.net/critic.think/tests/Ennis-Weir.Critic.Think.Essay.Test.pdf>
- Evendi, E., Retnawati, H., & Istiyono, E. (2023). Assessment of critical mathematical thinking skills: A literature review. *Beta: Jurnal Tadris Matematika*, 16(2), 85–97. <https://doi.org/10.20414/betajtm.v16i2.592>
- Facione, P. A. (1990). *Critical Thinking: A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction*. American Philosophical Association. [https://www.researchgate.net/publication/242279575\\_Critical\\_Thinking\\_A\\_State](https://www.researchgate.net/publication/242279575_Critical_Thinking_A_State)

ment of Expert Consensus for Purposes of Educational Assessment and Instruction

- Facione, P. A. (2020). Critical Thinking: What It Is and Why It Counts. *Insight Assessment*. <https://docslib.org/doc/8867387/critical-thinking-what-it-is-and-why-it-counts-2020-update-peter-a>
- Gallegos Chicaiza, M. P., Gallegos Chicaiza, S. X., Guerrero Vaca, A. E., Chacón Cachiguango, E. A., Rengel Herrera, M. del R., & Trujillo Zapata, A. M. (2024). Desarrollo del pensamiento crítico matemático mediante problemas contextualizados en educación primaria. *Código Científico Revista De Investigación*, 5(2), 455–472. <https://doi.org/10.55813/gaea/ccri/v5/n2/563>
- Gioria, A. (2015). Tipos de razonamientos estadísticos informales que se ponen en relación a través de una tarea de inferencia estadística informal. <http://hdl.handle.net/11185/1426>
- Güner, P., & Gökçe, S. (2021). Linking critical thinking disposition, cognitive flexibility and achievement: Math anxiety's mediating role. *The Journal of Educational Research*, 114(5), 458–473. <https://doi.org/10.1080/00220671.2021.1975618>
- Hang, N. V. T. (2020). Design of a competency-based moral lesson for teaching critical thinking skills in Vietnamese primary schools. *Issues In Educational Research*, 30(1), 133-155. <https://search.informit.org/doi/abs/10.3316/informit.085860750184238>
- Hernández-Sampieri, R., Fernández-Collado, C., y Baptista-Lucio, P. (2014). *Metodología de la investigación* (6.<sup>a</sup> ed.). McGraw-Hill. <https://dialnet.unirioja.es/servlet/libro?codigo=775008>
- Hernández-Sampieri, R., y Mendoza-Torres, C. (2023). *Metodología de la investigación. Las rutas cuantitativas, cualitativas y mixtas* (2.<sup>a</sup> ed.). McGraw-Hill.
- Imbo, I., & Vandierendonck, A. (2007). The development of strategy use in elementary school children: Working memory and individual differences. *Journal of Experimental Child Psychology*, 96(4), 284–309. <https://doi.org/10.1016/j.jecp.2006.09.001>
- Insight Assessment. (2007). *California Critical Thinking Disposition Inventory (CCTDI)*. <https://insightassessment.com/iaproduct/california-critical-thinking-disposition-inventory/>
- Insight Assessment. (2010). *California Critical Thinking Skills Test (CCTST)*. <https://insightassessment.com/iaproduct/california-critical-thinking-skills-test/>
- Insight Assessment. (s.f.). *K-12 Reasoning and Imagination Scales for Thinking and Learning (KRISTAL)*. <https://insightassessment.com/iaproduct/kristal/>

- International Test Commission. (2017). *The ITC Guidelines for Translating and Adapting Tests* (Second edition). [https://www.intestcom.org/files/guideline\\_test\\_adaptation\\_2ed.pdf](https://www.intestcom.org/files/guideline_test_adaptation_2ed.pdf)
- Jiang, R., Liu, R-d., Star, J., Zhen, R., Wang, J., Hong, W., Jiang, S., Sun, Y., & Fu, X. (2021). How mathematics anxiety affects students' inflexible perseverance in mathematics problem-solving: Examining the mediating role of cognitive reflection. *British Journal of Educational Psychology*, 91(1), 237–260. <https://doi.org/10.1111/bjep.12364>
- Jiménez Herrera, M. (2025). Guía práctica de validez y confiabilidad en instrumentos de evaluación para la investigación en salud. *Universidad del Desarrollo, Chile. Primera edición*. <https://medicina.udd.cl/innovatm/files/2025/09/validez-y-confiabilidad.pdf>
- Jorba, J., Gómez, I., y Prat, A. (2000). *Hablar y escribir para aprender.: Uso de la lengua en situación de enseñanza-aprendizaje desde las áreas curriculares*. Madrid: Editorial Síntesis. <https://archive.org/details/hablaryescribirp0000unse/page/5/mode/1up>
- Kettler, T. (2014). Critical Thinking Skills Among Elementary School Students: Comparing Identified Gifted and General Education Student Performance. *Gifted Child Quarterly*, 58(2), 127-136. <https://doi.org/10.1177/0016986214522508>
- Lai, E. R. (2011). Critical thinking: A literature review. *Pearson's research reports*, 6(1), 40-41. [https://www.researchgate.net/publication/297782058\\_Critical\\_thinking\\_A\\_literature\\_review](https://www.researchgate.net/publication/297782058_Critical_thinking_A_literature_review)
- Lin, I-H., Chew, S. W., & Chen, N-S. (2017). A vocal assessment approach to measure elementary school students' critical thinking skills. *IEEE 17th International Conference on Advanced Learning Technologies (ICALT)*, 419-421. <https://doi.org/10.1109/icalt.2017.30>
- Lipman, M. (1998). Pensamiento complejo y educación. *Ediciones de la Torre*. 10(2). <https://ia903101.us.archive.org/35/items/LipmanPensamientoComplejoYEducacion/Lipman%20-%20pensamiento%20complejo%20y%20educaci%C3%B3n.pdf>
- Lugo, J. G., García, J. I., y Ruíz, B. R. (2020). Desarrollo de la Investigación en Razonamiento Inferencial Informal. *Revista Digital: Matemática, Educación e Internet*, 20(1), 1-13. <https://www.redalyc.org/pdf/6079/607973099002.pdf>
- Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research*, 35(6), 382–385. <https://doi.org/10.1097/00006199-198611000-00017>

- Maričić, S., y Špijunović, K. (2015). Developing Critical Thinking in Elementary Mathematics Education through a Suitable Selection of Content and Overall Student Performance. *Procedia - Social and Behavioral Sciences*, 180, 653-659. <https://doi.org/10.1016/j.sbspro.2015.02.174>
- Massa, S. (2014). The development of critical thinking in primary school: the role of teacher's beliefs. *Procedia-Social and Behavioral Sciences*, 141(25), 387-392. <http://doi.org/10.1016/j.sbspro.2014.05.068>
- McNamara, D. S. (2004). Aprender del texto: Efectos de la estructura textual y las estrategias del lector. *Revista Signos*, 37 (55), 1-12. <https://www.redalyc.org/articulo.oa?id=157013765002>
- Meliá, J. L. (2001). *Teoría de la Fiabilidad y la Validez* 10 (3), 299-304. <https://www.uv.es/~meliajl/Research/LibroTFVpdf/TFVindex.htm>
- Ministerio de Educación de Chile. (2024). *Bases Curriculares: Primero a Sexto Básico*. [https://www.curriculumnacional.cl/614/articles-22394\\_bases.pdf](https://www.curriculumnacional.cl/614/articles-22394_bases.pdf)
- Muñiz, J. y Fonseca-Pedrero, E. (2019). Diez pasos para la construcción de un test. *Psicothema*, 31(1), 7-16. <https://doi.org/10.7334/psicothema2018.291>
- Nomen, J. (2019). La escuela, ¿un receptáculo del pensamiento crítico? *Folia Humanística*, (11). <https://doi.org/10.30860/0048>
- OECD. (2019). *Education at a Glance 2019: OECD Indicators*. OECD Publishing, Paris. <https://doi.org/10.1787/f8d7880d-en>.
- O'Reilly, C., Devitt, A., y Hayes, N. (2022). Critical thinking in the preschool classroom: A systematic literature review. *Early Childhood Research Quarterly*, 46(60), 1–15. <https://doi.org/10.1016/j.tsc.2022.101110>
- Ortega-Quevedo, V., Gil-Puente, C., Vallés-Rapp, C., y López-Luengo, M. (2020). Diseño y validación de instrumentos de evaluación de pensamiento crítico en educación primaria. *Tecné, Episteme y Didaxis: TED*, (48), 91-110. <https://doi.org/10.17227/ted.num48-12383>
- Ossa-Cornejo, C. J., Palma-Luengo, M. R., Lagos-San Martín, N. G., Quintana-Abello, I. M., y Díaz-Larenas, C. H. (2017). ANÁLISIS DE INSTRUMENTOS DE MEDICIÓN DEL PENSAMIENTO CRÍTICO. *Ciencias Psicológicas*, 11(1), 19-28. <https://doi.org/10.22235/cp.v11i2.1343>
- Palma, R. L. y Bores, R. E. (2002). *La importancia de desarrollar el pensamiento crítico en la escuela primaria para la obtención de aprendizajes significativos*. [Tesis de

- maestría, Instituto Tecnológico y de Estudios Superiores de Monterrey], México.  
<http://hdl.handle.net/11285/569921>.
- Penfield, R. D., y Giacobbi, Jr, P. R. (2004). Applying a score confidence interval to Aiken's item content-relevance index. *Measurement in physical education and exercise science*, 8(4), 213-225.  
[https://www.tandfonline.com/doi/abs/10.1207/s15327841mpee0804\\_3](https://www.tandfonline.com/doi/abs/10.1207/s15327841mpee0804_3)
- Pfannkuch, M. (2006). Informal inferential reasoning.  
[https://www.stat.auckland.ac.nz/~iase/publications/17/6A2\\_PFAN.pdf](https://www.stat.auckland.ac.nz/~iase/publications/17/6A2_PFAN.pdf)
- Ramirez, G., Hooper, S. Y., Kersting, N. B., Ferguson, R., & Yeager, D. (2018). Teacher math anxiety relates to adolescent students' math achievement. *AERA Open*, 4(1)  
<https://doi.org/10.1177/2332858418756052>
- Rodríguez, M. I. (2013). Inferencia informal: del análisis de los datos a la inferencia estadística. María Inés Rodríguez. *Revista de Educación Matemática*.  
<https://doi.org/10.33044/revem.10167>
- Rubin, A., Hammerman, J. y Konold, C. (2006). Exploración de la inferencia informal con software de visualización interactiva.  
<https://www.srri.umass.edu/sites/srri/files/rubin-2006eii/index.pdf>
- Smith-Castro, V., Molina-Delgado, M. (2011). Técnicas de entrevista cognitiva para la validación de cuestionarios. Instituto de Investigaciones Psicológicas, Universidad de Costa Rica. *Serie Cuadernos Metodológicos*, 5.  
[https://www.researchgate.net/publication/236335393\\_La\\_entrevista\\_cognitiva\\_Guia\\_para\\_su\\_aplicacion\\_en\\_la\\_evaluacion\\_y\\_mejoramiento\\_de\\_instrumentos\\_de\\_papel\\_y\\_lapiz](https://www.researchgate.net/publication/236335393_La_entrevista_cognitiva_Guia_para_su_aplicacion_en_la_evaluacion_y_mejoramiento_de_instrumentos_de_papel_y_lapiz)
- Smetanová, V., Drbalová, A. y Vitáková, D. (2015). Implicit theories of critical thinking in teachers and future teachers. *Procedia-Social and Behavioral Sciences*, 171(16), 724-732. <https://doi.org/10.1016/j.sbspro.2015.01.184>
- Sosa-Gutierrez, F., Vilca Apaza, H., Valdivia-Yábar, S., & Condori-Castillo, W. (2023). Critical thinking and teaching mathematics: An analysis from education. *International Journal of Religion*, 5(9), 958-976.  
<https://doi.org/10.61707/94v23344>
- Sparapani, N., Tseng, N., Towers, L., Birkeneder, S., Karimi, S., Alexander, C. J., Vega Garcia, J., Wood, T., & Dimachkie Nunnally, A. (2023). *Factores asociados a la participación en clase de alumnos de preescolar a tercer grado con espectro autista*. <https://doi.org/10.1016/j.rasd.2023.102186>
- Tamayo, O., Zona, R. y Loaiza, Y. (2015) El pensamiento crítico en la educación. Algunas categorías centrales en su estudio. *Revista Latinoamericana de Estudios*

<https://www.redalyc.org/pdf/1341/134146842006.pdf>

- The Critical Thinking Co. (2005). *Cornell Critical Thinking Tests*. <https://www.criticalthinking.com/cornell-critical-thinking-tests.html>
- The Critical Thinking Co. (2005). *Cornell Critical Thinking Test Level X*. <https://www.criticalthinking.com/cornell-critical-thinking-test-level-x.html>
- The Critical Thinking Co. (2005). *Cornell Critical Thinking Test Level Z*. <https://www.criticalthinking.com/cornell-critical-thinking-test-level-z.html>
- UNESCO. (2025). *Cultura y educación*. <https://www.unesco.org/es/cultura-y-educacion>
- Uribe-Hincapié, R. y Gutiérrez-Ríos, M. (2023). Concepciones acerca del pensamiento crítico en la enseñanza primaria y secundaria. Una revisión bibliográfica. *Revista Innova Educación*, 5(2), 91-107. <https://doi.org/10.35622/j.rie.2023.02.006>
- Vásquez, E. (2022). Modelo de estrategias de indagación para el desarrollo del pensamiento crítico y creativo en estudiantes de educación primaria. *Revista Innova Educación*, 4(3), 126–136. <https://doi.org/10.35622/j.rie.2022.03.008>
- Watson, G., y Glaser, E. M. (1980). *Watson-Glaser Critical Thinking Appraisal*. San Antonio, TX: Psychological Corporation.
- Willingham, D. T. (2008) Critical Thinking: Why Is It So Hard to Teach?, *Arts Education Policy Review*, 109(4), 21-32, <http://dx.doi.org/10.3200/AEPR.109.4.21-32>
- Wijayanti, R., Muhari, M., & Setyowati, R. (2018). Application of Problem Based Learning Model to Increase Learning Activities and Critical Thinking Skills. *Proceedings Of The 2nd International Conference On Education Innovation (ICEI)*. <https://doi.org/10.2991/icei-18.2018.12>
- Zaidi-Bibler, N., Grob, K., Monrad, S., Kurtz, J., Tai, A., Ahmed, A., Gruppen, L., y Santen, S. (2018). Pushing Critical Thinking Skills With Multiple-Choice Questions: Does Bloom's Taxonomy Work? *Acad Med*. 2018 Jun;93(6):856-859. <https://doi.org/10.1097/acm.000000000000208>

## BIBLIOGRAFÍA

- Attorresi, H. F., Lozzia, G. S., Abal, F. J., Galibert, M. S., y Aguerri, M. E. (2009). Teoría de Respuesta al Ítem. Conceptos básicos y aplicaciones para la medición de constructos psicológicos. *Revista Argentina de Clínica Psicológica*, 18(2), 179-188. <https://www.redalyc.org/articulo.oa?id=281921792007>
- Cronbach, L. J. y Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302. <https://psycnet.apa.org/doi/10.1037/h0040957>
- Hu, W., Adey, P., Jia, X., Liu, J., Zhang, L., Li, J. and Dong, X. (2011), Effects of a ‘Learn to Think’ intervention programme on primary school students. *British Journal of Educational Psychology*, 81, 531-557. <https://doi.org/10.1348/2044-8279.002007>
- Nunnally, J., y Bernstein, I. (1994). *Psychometric Theory* (3ª ed.). Nueva York: McGraw-Hill. <https://doi.org/10.1007/BF02301419>
- Popat, S., & Starkey, L. (2019). Learning to code or coding to learn? A systematic review. *Computers & Education*, 128, 365-376. <https://doi.org/10.1016/j.compedu.2018.10.005>
- Sternberg, R. J. (1986). Critical Thinking: Its Nature, Measurement, and Improvement. *National Inst. of Education (ED), Washington, DC*. <https://files.eric.ed.gov/fulltext/ED272882.pdf>
- Tripon, C. (Țîmpău). (2017). *Four key skills – reference framework for integrated curriculum*. The European Proceedings of Social and Behavioural Sciences (EpSBS), 518-525. <https://doi.org/10.15405/epsbs.2017.05.02.64>
- Warneken, F., y Tomasello, M. (2009). Varieties of altruism in children and chimpanzees. *Trends in Cognitive Sciences*, 13(9), 397-402. <http://dx.doi.org/10.1016/j.tics.2009.06.008>
- Whiteboard, D. y Basilio, M. (2012). Emergencia y desarrollo temprano de la autorregulación en niños preescolares. Profesorado. *Revista de Currículum y Formación de Profesorado*, 16(1), 15-34. <https://www.redalyc.org/pdf/567/56724377002.pdf>
- Zieffler, A., Garfield, J., DelMas, R. y Reading, C. (2008). Un marco para apoyar la investigación sobre razonamiento inferencial informal. *Revista de Investigación en Educación Estadística*, 7 (2), 40-58. <https://doi.org/10.52041/serj.v7i2.469>

## **ANEXOS**

### **8.1 ANEXO 1: INSTRUMENTO DE VALIDACIÓN**

#### **INSTRUMENTO DE VALIDACIÓN: PENSAMIENTO CRÍTICO EN CONTEXTOS DE RAZONAMIENTO INFERENCIAL INFORMAL (PC-RII)**

Estimados jueces,

Este trabajo se enmarca en el Fondecyt de iniciación N°11250799 “Explorando el desarrollo del pensamiento crítico y la curiosidad en matemática: un estudio transversal con estudiantes de primero a cuarto básico que participan en secuencias instruccionales diseñadas en el contexto del Estudio de Clases” dirigido por el Dr. Sergio Morales Candía y las tesis de la carrera de educación general básica Josselyn Águila, Bárbara Erices y Carolina Salgado. Desde ya, le agradecemos su participación en la validación de constructo de este instrumento.

El objetivo de este estudio es diseñar y validar un test psicométricamente robusto que permita medir el pensamiento crítico en estudiantes de 7 a 10 años en el área de matemática. Para ello, el instrumento será revisado por cinco jueces expertos, de acuerdo con las recomendaciones metodológicas de Lynn (1986), utilizando como criterios de valoración la claridad, pertinencia, contextualización, relevancia de las alternativas y cobertura de la evaluación del pensamiento crítico.

Es importante considerar que el instrumento se fundamenta en el modelo de pensamiento crítico descrito en el informe Informe Delphi (Facione, 1990; 2020), que reconoce seis habilidades principales: interpretación, análisis, evaluación, inferencia, explicación y autorregulación, cada una de ellas compuesta por sub-habilidades específicas. En consecuencia, los ítems han

sido elaborados para atender a cada una de las sub-habilidades específicas y adaptados al nivel cognitivo de los estudiantes de 7 a 10 años, procurando un lenguaje claro y cercano a su experiencia cotidiana.

De este modo, la revisión de los expertos resulta fundamental para fortalecer la validez de contenido del instrumento, asegurando que cada ítem mida efectivamente la sub-habilidad que se propone evaluar, y que su redacción sea adecuada para los niños y niñas que participarán en la investigación.

**Tabla 1.** Síntesis según concepto: pensamiento crítico, habilidades y sub-habilidades del instrumento.

<b>PENSAMIENTO CRÍTICO</b>			
El pensamiento crítico ha sido concebido como un proceso activo y reflexivo que guía la acción y la toma de decisiones (Dewey, 1910; Ennis, 2018), y como un proceso autocorrectivo, basado en criterios y consciente del contexto (Lipman, 1998). Para esta investigación, se adopta la definición de Facione (1990), Informe Delphi, que lo describe como un juicio autorregulado y con propósito, sustentado en habilidades cognitivas y disposiciones que permiten interpretar, analizar, evaluar e inferir de manera fundamentada (Facione, 2020).			
<b>HABILIDADES</b>	<b>Sub-habilidades</b>	<b>CANTIDAD DE ÍTEMS</b>	<b>CARACTERÍSTICAS GENERALES</b>
<b>A. Interpretación</b>	A1. Categorización	1	Comprender y comunicar el significado de información, experiencias o situaciones, integrando conocimientos científicos con perspectivas personales y sociales. Implica identificar, organizar y contextualizar datos o mensajes, reconociendo tanto sus significados explícitos como implícitos, para expresarlos de manera clara y coherente (Casas et al., 2005; Facione, 1990, Informe Delphi; Facione, 2020; Lin et al., 2017).
	A2. Decodificación del significado	1	
	A3. Clarificación del sentido	1	

<b>B. Análisis</b>	B1. Examen de ideas	1	Descomponer información, ideas o situaciones en sus elementos esenciales para comprender su lógica, evaluar razonamientos y reconocer relaciones inferenciales y conexiones entre conceptos. Facilita la identificación de premisas, conclusiones y suposiciones, así como la coherencia de los argumentos, promoviendo una comprensión profunda y crítica de la información (Elder y Paul, 2019; Facione, 1990, Informe Delphi; Facione, 2020; Lin et al., 2017).
	B2. Detección de argumentos	1	
	B3. Análisis de Argumentos	1	
<b>C. Evaluación</b>	C1. Valoración de la credibilidad	1	Juzgar la calidad del pensamiento propio y ajeno, valorando la credibilidad de la información y la solidez lógica de los argumentos. Implica identificar premisas, conclusiones y supuestos, anticipar objeciones y determinar la aceptabilidad de afirmaciones, contribuyendo a razonamientos sólidos y decisiones fundamentadas (Elder y Paul, 2019; Facione, 1990, Informe Delphi; Facione, 2020; Lin et al., 2017).
	C2. Valoración de los argumentos	1	
<b>D. Inferencia</b>	D1. Consulta de pruebas	1	Conectar información explícita e implícita con el conocimiento previo para extraer conclusiones fundamentadas. Implica formular hipótesis, generar alternativas y deducir consecuencias a partir de datos, principios o juicios, evaluando la plausibilidad de las opciones y fortaleciendo el razonamiento estratégico y la toma de decisiones basada en evidencia (Facione, 1990, Informe Delphi; Facione, 2020; McNamara, 2004; Lin et al., 2017).
	D2. Conjetura de alternativas	1	
	D3. Extracción de conclusiones	1	

<b>E. Explicación</b>	E1. Descripción de resultados	1	Presentar y comunicar de manera clara y fundamentada los resultados del razonamiento. Implica exponer relaciones causales, justificar procedimientos y ofrecer argumentos que respalden conclusiones, facilitando la comprensión de los hechos y la evaluación de su validez (Facione, 1990, Informe Delphi; Facione, 2020; Jorba et al., 2000; Lin et al., 2017).
	E2. Justificación de procedimientos	1	
	E3. Presentación de argumentos	1	
<b>F. Autorregulación</b>	F1. Autoexamen	1	Supervisar y ajustar el propio razonamiento, reconocer errores o prejuicios, y corregirlos mediante estrategias apropiadas. Implica reflexionar sobre juicios, conclusiones y procedimientos, así como controlar conductas y emociones para mantener un pensamiento objetivo, coherente y eficaz (Facione, 1990, Informe Delphi; Facione, 2020; Tamayo et al., 2015; Whitebread y Basilio, 2012).
	F2. Autocorrección.	1	

### **Sobre la validación**

La valoración del instrumento por ítem se organiza en cinco criterios: claridad, pertinencia en la medición de la sub-habilidad, contextualización para el razonamiento inferencial informal, relevancia de las alternativas y cobertura del pensamiento crítico por ítem, mientras que la valoración del instrumento general se organiza en un criterio: cobertura de la evaluación del pensamiento crítico. Cada categoría se evalúa mediante una escala Likert con cuatro niveles de desempeño (con puntajes de 1 a 4), la puntuación se asigna de la siguiente manera:

Nivel de desempeño	Deficiente	Regular	Bueno	Excelente
Puntaje	1 punto	2 puntos	3 puntos	4 puntos

Además, cada pauta incluye un espacio para observaciones cualitativas, de modo que el experto pueda aportar comentarios, sugerencias o ajustes necesarios desde su perspectiva.

La Tabla 2 presenta una síntesis de cada criterio, pregunta orientadora, su puntaje y el indicador correspondiente, que deben considerarse al momento de valorar cada ítem.

**Tabla 2.** Descripción de categorías, preguntas orientadoras y sus indicadores según el puntaje asignado.

Criterio	Pregunta orientadora	Puntaje	Indicador
<b>Valoración por ítem</b>			
<b>Claridad</b> El ítem es comprensible y utiliza un lenguaje claro y adecuado para niños de 7-10 años.	¿La redacción es clara y fácil de entender para niños de 7-10 años, sin ambigüedades?	1	El ítem no se comprende. La estructura y el vocabulario son inadecuados, impidiendo entender la consigna. Necesita rediseño total.
		2	El ítem se entiende con esfuerzo. Presenta ambigüedades o términos complejos que afectan su interpretación. Requiere reformulación.
		3	El ítem es claro en general, pero usa al menos un elemento (palabra o frase) que puede causar duda o que no es totalmente adecuado para la edad. Sugiere revisiones puntuales.
		4	El ítem es inmediatamente comprensible.

			El lenguaje y la estructura son precisos para la edad. No requiere modificaciones.
<b>Pertinencia</b> El ítem planteado es pertinente y está alineado con la habilidad y sub-habilidad que está midiendo.	¿El ítem evalúa específicamente la sub-habilidad declarada, sin incluir elementos irrelevantes o distractores?	1	El ítem no evalúa la sub-habilidad indicada. Su contenido es irrelevante o mide algo completamente distinto. Requiere un rediseño completo.
		2	El ítem evalúa la habilidad general pero no se enfoca en la sub-habilidad específica, o incluye elementos que distraen de ella. Necesita reestructuración para alinearse con la sub-habilidad.
		3	El ítem evalúa correctamente la sub-habilidad, pero presenta detalles menores (ej: redacción, formato, opciones) que podrían optimizarse para mayor precisión. Sugiere ajustes puntuales.
		4	El ítem mide exactamente la sub-habilidad definida, sin elementos irrelevantes o distractores. Es preciso y no requiere cambios.
<b>Contextualización para el razonamiento inferencial informal</b> Evalúa si el ítem presenta información, ejemplos, situaciones o gráficos significativos para niños de 7–10 años, que les	¿El ítem presenta un contexto significativo y adecuado a la edad, que facilite a los niños de 7–10 años realizar inferencias de manera informal, alineado con la habilidad y sub-	1	El ítem carece de contexto significativo o relevante para la edad. Los ejemplos, situaciones o gráficos no permiten realizar el razonamiento inferencial informal ni se relacionan con la habilidad o sub-habilidad indicada. Requiere rediseño completo para integrar un contexto que active el razonamiento inferencial informal.
		2	El ítem presenta algún grado de contextualización, pero los ejemplos, situaciones o gráficos son poco relevantes, no favorecen claramente el razonamiento inferencial informal o no se alinean del todo con la habilidad y sub-habilidad. Puede generar confusión o no activar el razonamiento inferencial de forma natural. Requiere reestructuración para mejorar la conexión, pertinencia e inferencia.

<p>permitan inferir conclusiones de manera indirecta (a partir de pistas o datos implícitos), conectados con su experiencia cotidiana y alineados con la habilidad y sub-habilidad que se busca medir.</p>	<p>habilidad indicadas?</p>	3	<p>El ítem es mayormente significativo y permite realizar el razonamiento inferencial informal relacionadas con la habilidad y sub-habilidad, pero algunos elementos podrían adaptarse mejor para reforzar el razonamiento inferencial informal.</p> <p>Sugiere ajustes puntuales para optimizar la comprensión y la capacidad de inferir.</p>
		4	<p>El ítem está completamente contextualizado, es significativo para la edad y activa de forma natural el razonamiento inferencial informal, alineado con la habilidad y sub-habilidad.</p> <p>Los ejemplos, situaciones o gráficos conectan claramente con la experiencia cotidiana y permiten deducir conclusiones de forma indirecta. No requiere cambios.</p>
<p><b>Relevancia de las alternativas</b> Las opciones de respuesta son claras, pertinentes y coherentes con el enunciado, evaluando la habilidad y sub-habilidad sin generar ambigüedades ni incluir distractores irrelevantes.</p>	<p>¿Las opciones de respuesta están directamente relacionadas con el enunciado y permiten medir con precisión la habilidad y sub-habilidad declaradas, sin incluir elementos irrelevantes o ambiguos?</p>	1	<p>Las opciones no evalúan la habilidad ni sub-habilidad indicada. Son irrelevantes, confusas o carecen de relación con el enunciado, lo que invalida la medición.</p> <p>Requiere rediseño total.</p>
		2	<p>Las opciones no integran de manera efectiva la habilidad y sub-habilidad con el enunciado.</p> <p>Algunas alternativas son confusas, irrelevantes o no se contextualizan adecuadamente con la situación planteada.</p> <p>Requiere reformular para asegurar coherencia y pertinencia.</p>
		3	<p>Las opciones evalúan adecuadamente la habilidad y sub-habilidad, y son mayormente coherentes con el enunciado.</p> <p>Presentan leves deficiencias (ejemplo: una opción descontextualizada o redacción mejorable).</p> <p>Sugiere ajustes puntuales para reforzar coherencia y claridad.</p>
		4	<p>Todas las opciones evalúan con precisión la habilidad y sub-habilidad, y son totalmente coherentes con el enunciado.</p> <p>No contienen distractores ni ambigüedades.</p>

			No requiere cambios.
<b>Cobertura del pensamiento crítico</b> Evalúa si el ítem mide de forma auténtica el pensamiento crítico, abordando de manera precisa la habilidad y sub-habilidad asignadas, asegurando que los componentes relevantes del constructo estén representados en la tarea propuesta.	¿Este ítem evalúa de forma auténtica el pensamiento crítico, alineándose con la habilidad y sub-habilidad que tiene asignadas?	1	El ítem no refleja el constructo de pensamiento crítico o solo aborda una parte mínima de él. No hay representación de la habilidad y sub-habilidad definida. Requiere rediseño total del ítem.
		2	El ítem incluye algunos elementos vinculados al pensamiento crítico, pero no desarrolla plenamente la habilidad y sub-habilidad asignadas. Presenta enfoque superficial o centrado en un único aspecto, dejando de lado componentes clave. Requiere reestructuración para alinearse con la habilidad y sub-habilidad.
		3	El ítem evalúa la habilidad y sub-habilidad asignadas y demanda pensamiento crítico en general, pero presenta aspectos que podrían optimizarse. Puede existir un leve desequilibrio o falta de profundidad en el planteamiento. Sugiere ajustes puntuales para reforzar la medición.
		4	El ítem evalúa de manera precisa y completa la habilidad y sub-habilidad asignadas, promoviendo un uso auténtico del pensamiento crítico. No presenta elementos irrelevantes ni distractores y está totalmente alineado con el constructo. No requiere cambios.

Agradeciendo de antemano su colaboración, se despiden atentamente:

Dr. Sergio Orlando Morales Candia, Universidad de Concepción, director Fondecyt de iniciación N°11250799

Josselyn Noelia Águila Ulloa, Bárbara Anaís Erices García, Carolina Angélica Salgado Salgado, tesistas del proyecto, Escuela de Educación, Universidad de Concepción.

### **EVALUACIÓN DE LA VALIDEZ DE CONTENIDO**

A continuación, se presentan los ítems organizados según su habilidad y sub-habilidad, junto con los criterios de valoración correspondientes.

Por favor:

1. Asigne el puntaje que considere adecuado para cada criterio en cada ítem.
2. Incluya comentarios específicos que permitan realizar los ajustes necesarios.

#### **Antecedentes del validador**

<b>Nombre</b>	
<b>Filiación institucional</b>	
<b>Estudios de posgrado</b>	
<b>Línea de investigación principal</b>	
<b>Experiencia en relación con la psicología educativa y didáctica de la matemática</b>	

## VALORACIÓN POR ÍTEM

<b>ÍTEM N° 1</b>					
<b>Código</b>	<b>Habilidad</b>			<b>Sub-habilidad</b>	
A1	<b>Interpretación</b>	Comprender y expresar el significado o importancia de una amplia variedad de experiencias, situaciones, datos, eventos, juicios, convenciones, creencias, reglas, procedimientos o criterios.		<b>Categorización</b>	Aprender a formular apropiadamente categorías, distinciones o marcos para comprender, describir o caracterizar la información. Describir experiencias, situaciones, creencias, acontecimientos, etc., de forma que adquieran significados comprensibles en términos de categorizaciones, distinciones o marcos apropiados.
<b>Puntaje por criterio</b>					
<b>Claridad</b>	<b>Pertinencia</b>	<b>Contextualización para el razonamiento inferencial informal</b>	<b>Relevancia de las alternativas</b>	<b>Cobertura del pensamiento crítico por ítem</b>	
[Si considera necesario, incluya comentarios adicionales aquí]					

<b>ÍTEM N° 2</b>					
<b>Código</b>	<b>Habilidad</b>			<b>Sub-habilidad</b>	
A2	<b>Interpretación</b>	Comprender y expresar el significado o importancia de una		<b>Decodificación del significado</b>	Detectar, atender y describir el contenido informativo, el

		amplia variedad de experiencias, situaciones, datos, eventos, juicios, convenciones, creencias, reglas, procedimientos o criterios.		significado afectivo, las funciones directivas, intenciones, motivos, propósitos, importancia social, valores, puntos de vista, reglas, procedimientos, criterios o relaciones inferenciales expresados en sistemas de comunicación basados en convenciones, tales como lenguaje, conductas sociales, dibujos, números, gráficos, tablas, cuadros, signos y símbolos.
<b>Puntaje por criterio</b>				
<b>Claridad</b>	<b>Pertinencia</b>	<b>Contextualización para el razonamiento inferencial informal</b>	<b>Relevancia de las alternativas</b>	<b>Cobertura del pensamiento crítico por ítem</b>
<b>[Si considera necesario, incluya comentarios adicionales aquí]</b>				

<b>ÍTEM N° 7</b>				
<b>Código</b>	<b>Habilidad</b>		<b>Sub-habilidad</b>	
A3	<b>Interpretación</b>	Comprender y expresar el significado o importancia de una amplia variedad de experiencias, situaciones, datos,	<b>Clarificación del sentido</b>	Parafrasear o hacer explícitos, mediante estipulación, descripción, analogía o expresión figurativa, los

		eventos, juicios, convenciones, creencias, reglas, procedimientos o criterios.		significados contextuales, convencionales o intencionados de palabras, ideas, conceptos, declaraciones, comportamiento, dibujos, números, signos, cuadros, gráficos, símbolos, reglas, eventos o ceremonias. Usar estipulación, descripción, analogía o expresión figurativa para eliminar confusión, vaguedad o ambigüedad no intencionada, o diseñar un procedimiento razonable para hacerlo.
<b>Puntaje por criterio</b>				
<b>Claridad</b>	<b>Pertinencia</b>	<b>Contextualización para el razonamiento inferencial informal</b>	<b>Relevancia de las alternativas</b>	<b>Cobertura del pensamiento crítico por ítem</b>
<b>[Si considera necesario, incluya comentarios adicionales aquí]</b>				

ÍTEM N° 5				
Código	Habilidad		Sub-habilidad	
B1	<b>Análisis</b>	Identificar las relaciones inferenciales previstas y reales entre afirmaciones, preguntas, conceptos, descripciones u otras	<b>Examen de ideas</b>	Determinar el papel que desempeñan o pretenden desempeñar diversas expresiones en el contexto de

		formas de representación destinadas a expresar creencias, juicios, experiencias, razones, información u opiniones.		argumentación, razonamiento o persuasión. Definir términos. Comparar o contrastar ideas, conceptos o afirmaciones. Identificar cuestiones o problemas y determinar sus partes componentes y también identificar las relaciones conceptuales de esas partes entre sí y con el todo.
<b>Puntaje por criterio</b>				
<b>Claridad</b>	<b>Pertinencia</b>	<b>Contextualización para el razonamiento inferencial informal</b>	<b>Relevancia de las alternativas</b>	<b>Cobertura del pensamiento crítico</b>
<b>[Si considera necesario, incluya comentarios adicionales aquí]</b>				

<b>ÍTEM N° 9</b>				
<b>Código</b>	<b>Habilidad</b>		<b>Sub-habilidad</b>	
B2	<b>Análisis</b>	Identificar las relaciones inferenciales previstas y reales entre afirmaciones, preguntas, conceptos, descripciones u otras formas de representación destinadas a expresar creencias, juicios, experiencias, razones, información u opiniones.	<b>Detección de argumentos</b>	Dado un conjunto de afirmaciones, descripciones, preguntas o representaciones gráficas, determinar si el conjunto expresa, o pretende expresar, una o más razones en favor o

				en contra de alguna afirmación, opinión o punto de vista.
<b>Puntaje por criterio</b>				
<b>Claridad</b>	<b>Pertinencia</b>	<b>Contextualización para el razonamiento inferencial informal</b>	<b>Relevancia de las alternativas</b>	<b>Cobertura del pensamiento crítico por ítem</b>
<b>[Si considera necesario, incluya comentarios adicionales aquí]</b>				

<b>ÍTEM N° 13</b>				
<b>Código</b>	<b>Habilidad</b>		<b>Sub-habilidad</b>	
B3	<b>Análisis</b>	Identificar las relaciones inferenciales previstas y reales entre afirmaciones, preguntas, conceptos, descripciones u otras formas de representación destinadas a expresar creencias, juicios, experiencias, razones, información u opiniones.	<b>Análisis de argumentos</b>	Dada la expresión de una o varias razones destinadas a apoyar o refutar alguna afirmación, opinión o punto de vista, para identificar y diferenciar: (a) la conclusión principal pretendida, (b) las premisas y razones presentadas en apoyo de la conclusión principal, (c) otras premisas y razones presentadas como respaldo o apoyo de esas premisas y razones destinadas a apoyar la conclusión principal, (d) elementos adicionales no expresados de ese razonamiento, tales como conclusiones intermedias, suposiciones o

				presuposiciones no declaradas, (e) la estructura general del argumento o la cadena de razonamiento pretendida, y (f) cualquier elemento contenido en el cuerpo de expresiones que se examina que no esté destinado a ser tomado como parte del razonamiento expresado o de su trasfondo previsto.
<b>Puntaje por criterio</b>				
<b>Claridad</b>	<b>Pertinencia</b>	<b>Contextualización para el razonamiento inferencial informal</b>	<b>Relevancia de las alternativas</b>	<b>Cobertura del pensamiento crítico por ítem</b>
[Si considera necesario, incluya comentarios adicionales aquí]				

<b>ÍTEM N° 4</b>				
<b>Código</b>	<b>Habilidad</b>		<b>Sub-habilidad</b>	
C1	<b>Evaluación</b>	Valorar la credibilidad de afirmaciones u otras representaciones que sean relatos o descripciones de la percepción, experiencia, situación, juicio, creencia u opinión de una persona; y valorar la fuerza lógica de las relaciones inferenciales reales o	<b>Valoración de la credibilidad</b>	Reconocer los factores relevantes para valorar el grado de credibilidad que se debe atribuir a una fuente de información u opinión. Valorar la relevancia contextual de preguntas, información, principios, reglas o directrices procedimentales.

		intencionadas entre afirmaciones, descripciones, preguntas u otras formas de representación.		Valorar la aceptabilidad, el nivel de confianza que se debe depositar en la probabilidad o veracidad de cualquier representación dada de una experiencia, situación, juicio, creencia u opinión.
<b>Puntaje por criterio</b>				
<b>Claridad</b>	<b>Pertinencia</b>	<b>Contextualización para el razonamiento inferencial informal</b>	<b>Relevancia de las alternativas</b>	<b>Cobertura del pensamiento crítico por ítem</b>
<b>[Si considera necesario, incluya comentarios adicionales aquí]</b>				

<b>ÍTEM N° 10</b>				
<b>Código</b>	<b>Habilidad</b>		<b>Sub-habilidad</b>	
C2	<b>Evaluación</b>	Valorar la credibilidad de afirmaciones u otras representaciones que sean relatos o descripciones de la percepción, experiencia, situación, juicio, creencia u opinión de una persona; y valorar la fuerza lógica de las relaciones inferenciales reales o intencionadas entre afirmaciones, descripciones, preguntas u otras formas de representación.	<b>Valoración de los argumentos</b>	Juzgar si la aceptabilidad asumida de las premisas de un argumento dado justifica aceptar como verdadera (con certeza deductiva) o como muy probablemente verdadera (justificada inductivamente) la conclusión expresada en ese argumento. Anticipar o plantear preguntas u objeciones, y valorar si éstas apuntan a una debilidad significativa en el argumento que se está evaluando.

				<p>Determinar si un argumento se basa en suposiciones falsas o dudosas o presuposiciones y luego determinar cuán crucialmente éstas afectan su fortaleza.</p> <p>Juzgar entre inferencias razonables y falaces.</p> <p>Juzgar la fuerza probatoria de las premisas y supuestos de un argumento con miras a determinar la aceptabilidad del argumento.</p> <p>Determinar y juzgar la fuerza probatoria de un argumento intencionado o consecuencias no deseadas con vistas a juzgar la aceptabilidad del argumento.</p> <p>Determinar en qué medida una posible información adicional podría reforzar o debilitar un argumento.</p>
<b>Puntaje por criterio</b>				
<b>Claridad</b>	<b>Pertinencia</b>	<b>Contextualización para el razonamiento inferencial informal</b>	<b>Relevancia de las alternativas</b>	<b>Cobertura del pensamiento crítico por ítem</b>
[Si considera necesario, incluya comentarios adicionales aquí]				

ÍTEM N° 3				
Código	Habilidad		Sub-habilidad	
D1	<b>Inferencia</b>	Identificar y asegurar los elementos necesarios para extraer conclusiones razonables; formular conjeturas e hipótesis; considerar la información relevante y deducir las consecuencias que surgen de datos, afirmaciones, principios, evidencias, juicios, creencias, opiniones, conceptos, descripciones, preguntas u otras formas de representación.	<b>Consulta de pruebas</b>	En particular, reconocer las premisas que requieren apoyo y formular una estrategia para buscar y recopilar información que pueda proporcionar ese apoyo.  En general, para juzgar que la información es relevante para decidir la aceptabilidad. Se requiere la plausibilidad o los méritos relativos de una alternativa, pregunta, problema, teoría, hipótesis o afirmación dada, y determinar estrategias de investigación plausibles para adquirir esa información.
Puntaje por criterio				
Claridad	Pertinencia	Contextualización para el razonamiento inferencial informal	Relevancia de las alternativas	Cobertura del pensamiento crítico por ítem
[Si considera necesario, incluya comentarios adicionales aquí]				

ÍTEM N° 11				
Código	Habilidad		Sub-habilidad	
D2	<b>Inferencia</b>	Identificar y asegurar los elementos necesarios para extraer conclusiones razonables; formular conjeturas e hipótesis; considerar la información relevante y deducir las consecuencias que surgen de datos, afirmaciones, principios, evidencias, juicios, creencias, opiniones, conceptos, descripciones, preguntas u otras formas de representación.	<b>Conjetura de alternativas</b>	Formular múltiples alternativas para resolver un problema, postular una serie de suposiciones sobre una pregunta, proyectar hipótesis alternativas respecto a un acontecimiento, desarrollar una variedad de planes diferentes para alcanzar algún objetivo. Extraer presuposiciones y proyectar el rango de posibles consecuencias de decisiones, posiciones, políticas, teorías o creencias.
Puntaje por criterio				
Claridad	Pertinencia	Contextualización para el razonamiento inferencial informal	Relevancia de las alternativas	Cobertura del pensamiento crítico por ítem
[Si considera necesario, incluya comentarios adicionales aquí]				

ÍTEM N° 8				
Código	Habilidad		Sub-habilidad	
D3	<b>Inferencia</b>	Identificar y asegurar los elementos necesarios para extraer conclusiones	<b>Extracción de conclusiones</b>	Aplicar modos apropiados de inferencia para determinar qué

		<p>razonables; formular conjeturas e hipótesis; considerar la información relevante y deducir las consecuencias que surgen de datos, afirmaciones, principios, evidencias, juicios, creencias, opiniones, conceptos, descripciones, preguntas u otras formas de representación.</p>	<p>posición, opinión o punto de vista que se debe adoptar sobre un asunto o cuestión determinada.</p> <p>Dado un conjunto de afirmaciones, descripciones, preguntas u otras formas de representación, deducir, con el nivel adecuado de fuerza lógica, sus relaciones inferenciales y las consecuencias o presuposiciones que ellas sustentan, garantizan, implican o conllevan.</p> <p>Emplear con éxito varias subespecies de razonamiento, como por ejemplo razonar analógicamente, aritméticamente, dialécticamente, científicamente, etc.</p> <p>Determinar cuál de varias conclusiones está más firmemente justificada o respaldadas por la evidencia disponible, o que deberían ser rechazadas o consideradas menos plausibles en función de la información proporcionada.</p>
<b>Puntaje por criterio</b>			

Claridad	Pertinencia	Contextualización para el razonamiento inferencial informal	Relevancia de las alternativas	Cobertura del pensamiento crítico por ítem
[Si considera necesario, incluya comentarios adicionales aquí]				

ÍTEM N° 12				
Código	Habilidad		Sub-habilidad	
E1	<b>Explicación</b>	Exponer los resultados del propio razonamiento; justificar ese razonamiento en términos de consideraciones evidenciables, conceptuales, metodológicas, criterio lógicas y contextuales.	<b>Descripción de resultados</b>	Producir declaraciones, descripciones o representaciones precisas de los resultados de las actividades de razonamiento de uno con el fin de analizar, evaluar, inferir o monitorear esos resultados.
<b>Puntaje por criterio</b>				
Claridad	Pertinencia	Contextualización para el razonamiento inferencial informal	Relevancia de las alternativas	Cobertura del pensamiento crítico por ítem
[Si considera necesario, incluya comentarios adicionales aquí]				

<b>ÍTEM N° 6</b>				
<b>Código</b>	<b>Habilidad</b>		<b>Sub-habilidad</b>	
E2	<b>Explicación</b>	Exponer los resultados del propio razonamiento; justificar ese razonamiento en términos de consideraciones evidenciables, conceptuales, metodológicas, criterio lógicas y contextuales.	<b>Justificación de procedimientos</b>	Presentar el análisis evidencial, conceptual, metodológico, criterio lógico y contextuales que se utilizan para formar sus interpretaciones, análisis, evaluaciones o inferencias, de modo que pueda registrar, evaluar, describir o justificar con precisión esos procesos para sí mismo o para otros, o para remediar deficiencias percibidas en la forma general en que uno ejecuta estos procesos.
<b>Puntaje por criterio</b>				
<b>Claridad</b>	<b>Pertinencia</b>	<b>Contextualización para el razonamiento inferencial informal</b>	<b>Relevancia de las alternativas</b>	<b>Cobertura del pensamiento crítico por ítem</b>
<b>[Si considera necesario, incluya comentarios adicionales aquí]</b>				

ÍTEM N° 14				
Código	Habilidad		Sub-habilidad	
E3	<b>Explicación</b>	Exponer los resultados del propio razonamiento; justificar ese razonamiento en términos de consideraciones evidenciables, conceptuales, metodológicas, criterio lógicas y contextuales.	<b>Presentación de argumentos</b>	Dar razones para aceptar alguna afirmación. Para hacer frente a objeciones al método, conceptualizaciones, evidencias, criterios o adecuación contextual de los juicios inferenciales, analíticos o evaluativos.
Puntaje por criterio				
	<b>Claridad</b>	<b>Pertinencia</b>	<b>Contextualización para el razonamiento inferencial informal</b>	<b>Relevancia de las alternativas</b>
				<b>Cobertura del pensamiento crítico por ítem</b>
[Si considera necesario, incluya comentarios adicionales aquí]				

ÍTEM N° 15				
Código	Habilidad		Sub-habilidad	
F1	<b>Autorregulación</b>	Supervisar conscientemente las propias actividades cognitivas, los elementos utilizados en esas actividades y los resultados obtenidos, en particular aplicando habilidades de análisis y evaluación	<b>Autoexamen</b>	Reflexionar sobre el propio razonamiento y verificar tanto los resultados producidos como la correcta aplicación y ejecución de las habilidades cognitivas implicadas.

		a los propios juicios inferenciales con vistas a cuestionar, confirmar, validar o corregir el propio razonamiento o los propios resultados.		<p>Hacer una autoevaluación metacognitiva objetiva y reflexiva de las propias opiniones y razones para sostenerlas.</p> <p>Juzgar hasta qué punto el pensamiento propio está influenciado por deficiencias en el propio conocimiento, o por estereotipos, prejuicios, emociones o cualquier otro factor que limite la objetividad o racionalidad de uno.</p> <p>Reflexionar sobre las propias motivaciones valores, actitudes e intereses con vistas a determinar que uno se ha esforzado por ser imparcial, justo, minucioso, objetivo, respetuoso de la verdad, razonable y racional al llegar a sus análisis, interpretaciones, evaluaciones, inferencias o expresiones.</p>
<b>Puntaje por criterio</b>				
<b>Claridad</b>	<b>Pertinencia</b>	<b>Contextualización para el razonamiento inferencial informal</b>	<b>Relevancia de las alternativas</b>	<b>Cobertura del pensamiento crítico por ítem</b>

[Si considera necesario, incluya comentarios adicionales aquí]

ÍTEM N° 16				
Código	Habilidad		Sub-habilidad	
F2	<b>Autorregulación</b>	Supervisar conscientemente las propias actividades cognitivas, los elementos utilizados en esas actividades y los resultados obtenidos, en particular aplicando habilidades de análisis y evaluación a los propios juicios inferenciales con vistas a cuestionar, confirmar, validar o corregir el propio razonamiento o los propios resultados.	<b>Autoexamen</b>	Cuando el autoexamen revele errores o deficiencias, diseñar medidas razonables, procedimientos para remediar o corregir, si es posible, dichos errores y sus causas.
Puntaje por criterio				
Claridad	Pertinencia	Contextualización para el razonamiento inferencial informal	Relevancia de las alternativas	Cobertura del pensamiento crítico por ítem
[Si considera necesario, incluya comentarios adicionales aquí]				

## 8.2 ANEXO 2: INSTRUMENTO PARA LA EVALUACIÓN DE ENTREVISTA COGNITIVA

### INSTRUMENTO PARA LA EVALUACIÓN DE ENTREVISTA COGNITIVA – PENSAMIENTO CRÍTICO EN RAZONAMIENTO INFERENCIAL INFORMAL (PC-RII)

#### I. INFORMACIÓN GENERAL

Nombre del estudiante: \_\_\_\_\_ Edad: \_\_\_\_\_ Curso: \_\_\_\_\_

Fecha de entrevista: \_\_\_\_\_ Entrevistador: \_\_\_\_\_

#### II. RÚBRICA DE EVALUACIÓN

Durante la resolución de cada ítem, el entrevistador debe realizar preguntas de sondeo protocolarias (ejemplo: “¿*Qué fue lo más difícil de entender?*”, “¿*Hay alguna palabra que no hayas comprendido?*”). En base a las respuestas verbales y a la observación de la conducta del estudiante, complete la rúbrica de evaluación correspondiente a ese ítem.

Cada criterio cuenta con cuatro indicadores observables. La puntuación en la escala Likert (1–5) se asigna de la siguiente manera:

1 punto	2 puntos	3 puntos	4 puntos	5 puntos
El ítem no cumple en absoluto con los indicadores. El ítem debe ser rediseñado completamente.	El ítem cumple deficientemente con los indicadores. El ítem requiere un ajuste estructural.	El ítem cumple de manera aceptable pero no óptima con los indicadores. El ítem requiere ajustes importantes.	El ítem cumple bien con los indicadores. La evidencia es mayoritariamente positiva. El ítem requiere menores ajustes.	El ítem cumple excelentemente. La evidencia es clara y fuerte en todos los aspectos (p. ej., los 4 indicadores son abordados de manera sobresaliente). El ítem

				no requiere modificaciones
--	--	--	--	----------------------------

En la columna de “Comentarios cualitativos” registre evidencia observable y ejemplos concretos que justifiquen la puntuación asignada.

En el apartado de “Sugerencias de mejora”, anote ajustes específicos que se recomienden para el ítem (ejemplo: “Sustituir la palabra ‘inferencia’ por ‘conclusión’”, “Reducir la longitud del enunciado”, “Evitar que dos opciones de respuesta sean demasiado similares”).

<b>Criterio a evaluar</b>	<b>Indicadores observables</b>	<b>Escala Likert (1 – 5)</b>	<b>Comentarios cualitativos</b>
<b>Comprensión semántica y lingüística</b>	<ul style="list-style-type: none"> <li>● El niño no pregunta el significado de ninguna palabra.</li> <li>● El niño puede parafrasear el enunciado usando sus propias palabras sin distorsionar el significado.</li> <li>● Su expresión facial no muestra confusión por términos clave.</li> <li>● No necesita que se le repita la pregunta de forma adicional.</li> </ul>		P. ej., “el estudiante frunce el ceño al leer la palabra inferencia.”
<b>Comprensión de la tarea y estructura</b>	<ul style="list-style-type: none"> <li>● Comprende que debe seleccionar una sola opción de respuesta.</li> <li>● Comprende cómo marcar su respuesta sin ayuda.</li> <li>● Entiende la instrucción general de la tarea.</li> </ul>		P. ej., “requiere que se le explique nuevamente cómo marcar la respuesta.”

	<ul style="list-style-type: none"> <li>● El formato de selección múltiple (a-b-c-d) le es familiar y manejable.</li> </ul>		
<b>Comprensión conceptual (Validez de constructo)</b>	<ul style="list-style-type: none"> <li>● Su explicación verbal del ítem se alinea con la habilidad de pensamiento crítico evaluada.</li> <li>● La alternativa elegida se ajusta con base en un razonamiento lógico, evidencial o comparativo.</li> <li>● No se distrae con elementos accesorios o irrelevantes del ítem.</li> <li>● Desacredita críticamente las alternativas no seleccionadas.</li> </ul>		P. ej., “Se distrae con el dibujo de la botella y menciona colores.”
<b>Engagement y pertinencia percibida</b>	<ul style="list-style-type: none"> <li>● Mantiene la atención durante la lectura y resolución del ítem.</li> <li>● Reconoce que la situación presentada le es familiar o posible.</li> <li>● Muestra interés genuino por el contenido de la pregunta.</li> <li>● El tiempo que tarda en responder es apropiado (ni inmediato, ni excesivamente largo por distracción).</li> </ul>		P. ej. “El estudiante demora 2:48 min en responder el ítem.”
<b>SUGERENCIAS DE MEJORA</b>			

### **8.3 ANEXO 3: MANUAL DE PROTOCOLO PARA LA APLICACIÓN DE LA ENTREVISTA COGNITIVA**

#### **MANUAL DE PROTOCOLO PARA LA APLICACIÓN DE LA ENTREVISTA COGNITIVA: EVALUACIÓN DEL PENSAMIENTO CRÍTICO EN RAZONAMIENTO INFERENCIAL INFORMAL**

Este manual establece el protocolo estandarizado para la aplicación de la entrevista cognitiva basada en la "Rúbrica de Entrevista Cognitiva".

Su objetivo es garantizar la validez, confiabilidad y equidad del proceso de evaluación, minimizando los sesgos del entrevistador y obteniendo datos cualitativos robustos sobre la comprensión y el procesamiento mental de los estudiantes.

#### **Objetivos específicos:**

- A. Estandarizar la interacción entre entrevistador y estudiante.
- B. Asegurar que las preguntas de sondeo se realicen de manera consistente.
- C. Guiar la observación y el registro de conductas observables.
- D. Uniformizar el proceso de calificación y generación de sugerencias de mejora para los ítems.

### **I. PREPARACIÓN PRE-ENTREVISTA**

#### **1.1. Logística y ambiente:**

Realice la entrevista en un espacio tranquilo, bien iluminado y libre de distracciones.

Tenga a mano:

- Una copia impresa de la Rúbrica de Evaluación.
- El instrumento (prueba o cuestionario) que el estudiante resolverá.
- Bolígrafo o lápiz para tomar notas.

- Un cronómetro (discreto) para medir el tiempo de respuesta.

### **1.2. Revisión del instrumento:**

El entrevistador debe estar totalmente familiarizado con cada ítem del instrumento que el estudiante va a resolver, incluyendo las instrucciones generales y las opciones de respuesta.

### **1.3. Presentación con el estudiante:**

Preséntese de manera cordial. Diga: "Hola [Nombre del estudiante], soy [Su nombre]. Hoy vamos a trabajar juntos en algunas preguntas. No es una prueba, sino que me ayudará a comprender cómo piensas"

Explique al estudiante que mientras resuelven los ítems, usted irá realizando preguntas de sondeo.

Asegure al estudiante que sus respuestas son confidenciales y se usarán solo para mejorar los materiales.

## **II. PROTOCOLO PARA LA APLICACIÓN DE ENTREVISTA COGNITIVA**

### **2.1. Flujo de la entrevista:**

- 1) Entregue el instrumento al estudiante.
- 2) Pídale que lea el ítem en voz alta.
- 3) Permita que el estudiante resuelva el ítem a su propio ritmo. Observe silenciosamente su conducta durante este proceso (expresiones faciales, gestos de confusión, tiempo de respuesta).
- 4) Inmediatamente después de que el estudiante marque su respuesta, proceda con las preguntas de sondeo, siguiendo el orden de los criterios de la rúbrica.

### **2.2. Preguntas de sondeo protocolarias por criterio:**

### **Fase 1: Evaluar la comprensión semántica y lingüística**

Verificar que el estudiante entienda el significado literal de las palabras y el enunciado.

- "¿Hay alguna palabra o frase en esta pregunta que no hayas entendido?"
- "¿Entendiste la pregunta? ¿necesitas que te la lea?"
- "¿Puedes explicarme con tus propias palabras de qué se trata el ejercicio?" (Busca una paráfrasis).

### **Fase 2: Evaluar la comprensión de la tarea y estructura**

Asegurarse de que el estudiante sabe QUÉ tiene que hacer y CÓMO hacerlo.

- "¿Qué es lo que tienes que hacer?"

### **Fase 3: Evaluar la comprensión conceptual (validez de constructo)**

Profundizar en el razonamiento para confirmar que la respuesta se basó en la habilidad de pensamiento crítico que se pretende evaluar.

- "¿Qué te hizo pensar que esa era la respuesta correcta?"
- "¿Por qué descartaste las otras opciones?"

### **Fase 4: Evaluar el engagement y pertinencia percibida**

Capturar la experiencia subjetiva del estudiante y su nivel de conexión con la tarea.

- "¿Qué te pareció el tema de la tarea?"
- La situación que se presenta o algo parecido ¿la has vivido en algún momento?

**NO:** ¿crees que es posible que pase algo así en la vida real?

**SI:** ¿en qué momento la viviste?

### **2.3. Técnica de observación y registro:**

Durante la resolución y el sondeo, observe activamente las señales no verbales listadas en la rúbrica (expresión facial de confusión, seguridad, distracción).

Tome notas de manera discreta pero continua en la columna de "Comentarios cualitativos". Anote frases literales del estudiante y descripciones concretas de su conducta.

- **Ejemplo de buen registro:** *"Al leer la palabra 'inferencia', frunció el ceño. Dijo: 'esta palabra no la he visto mucho'. Parafraseó el enunciado correctamente, pero omitió el concepto clave."*
- **Ejemplo de mal registro:** *"No entendió"* (Es vago y no sirve como evidencia).

### **III. PROTOCOLO DE CALIFICACIÓN POST-ENTREVISTA**

#### **3.1. Procedimiento de calificación:**

- 1) La calificación se realiza inmediatamente después de finalizar la entrevista con el estudiante.
- 2) Para cada ítem, revise sus notas cualitativas ("Comentarios").
- 3) Por cada criterio, cuente cuántos de los 4 indicadores observables se cumplieron.
- 4) Asigne la puntuación Likert (1-5) basándose exclusivamente en el conteo de indicadores cumplidos, tal como define la tabla de la rúbrica.

#### **3.2. Generación de sugerencias de mejora:**

Basándose en los problemas específicos observados (ej: una palabra incomprensible, una instrucción ambigua), redacte sugerencias accionables en el apartado correspondiente.

#### **3.3. Interpretación de resultados:**

Utilice la "Guía de Interpretación" para tomar decisiones sobre cada ítem.

### **IV. Principios éticos y buenas prácticas**

Mantenga una actitud neutral. No dé pistas, ni muestre aprobación o desaprobación con gestos ante las respuestas del estudiante.

Agradezca la participación con frases como *"Perfecto, esa explicación es justo lo que necesito"* o *"Gracias por ser tan detallado"*. Esto reduce la ansiedad.

Recuerde al estudiante (y a usted mismo) que el objetivo NO es evaluar su inteligencia, sino la claridad de las preguntas.

Asegure el anonimato del estudiante en todos los registros y reportes.