



UNIVERSIDAD DE CONCEPCIÓN
ESCUELA DE EDUCACIÓN

**DISEÑO Y VALIDACIÓN DE UNA ESCALA PARA MEDIR
LA CURIOSIDAD EPISTÉMICA EN MATEMÁTICAS EN
ESTUDIANTES DE SEGUNDO A CUARTO BÁSICO DEL
SISTEMA ESCOLAR CHILENO**

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

Autoras : Valeria Antonia Albial Henríquez

Guislene Stefania Cifuentes Labra

Profesor guía : Dr. Sergio Orlando Morales Candía

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, Valeria Antonia Albial Henríquez, Guislene Stefanía Cifuentes Labra.

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

Agradecimientos Guislene Stefania Cifuentes Labra.

En primer lugar, quiero expresar mi más sincero agradecimiento a mi compañera y amiga, Valeria. Iniciamos juntas esta etapa y ha sido un honor culminarla a tu lado. Gracias por confiar siempre en mí, por tu lealtad siempre, por tus ánimos en los momentos cruciales y por tu ayuda incondicional a lo largo de toda la carrera. Te quiero mucho y espero, de todo corazón, que nuestra amistad perdure en el tiempo.

Deseo extender un agradecimiento al profesor Sergio Morales por aceptar guiar este trabajo y por su constante disponibilidad. Su orientación fue fundamental para dar rigor y claridad a esta investigación, gracias por creer en nosotras y siempre impulsarnos a mirar más allá, seguir perfeccionándonos y trabajando en la investigación.

A mi familia, pilar fundamental de mi vida, mi madre, mi padre y mis hermanos. Su apoyo incondicional, presente antes, durante y al término de esta carrera, ha sido la base sobre la que construí cada paso. El respaldo de la familia es esencial en este proceso de forjar el futuro, y gracias a ustedes lo logré, superando incluso mis propias expectativas. Los amo profundamente; son parte de cada logro que he tenido y de todos los que vendrán.

Por último, pero no menos importante, agradezco a mi pareja Vicente Ramírez, quien ha estado a mi lado desde siempre. Tu apoyo incondicional fue primordial para mí, me consolaste en las dificultades, alegraste mis triunfos, me diste ánimo y siempre estuviste dispuesto a escuchar cada exposición o planificación, ofreciéndome tu opinión y colaborando en mi proceso de formación como futura docente. Mis logros son también tuyos, porque sin ti todo hubiera sido mucho más difícil. Te amo mucho. Ayer fue por mí, hoy es por ti y mañana será por nosotros.

Este título no es solo mío; es el reflejo del amor, la paciencia y la confianza de todos los que están en estas páginas. A cada uno de ustedes: gracias por ser parte de mi formación profesional.

Agradecimientos Valeria Antonia Albial Henríquez.

En primer lugar, quiero agradecer a mi compañera y amiga Guislene en esta etapa universitaria, por ser una amiga incondicional, autentica y leal. Mi colega y mi *motorcito*, gracias por esta amistad que nació para la vida.

Extiendo mi profundo agradecimiento a mi profesor guía, por creer en nosotras desde el primer momento, por su generosidad, paciencia y por todas las oportunidades brindadas.

Quiero expresar mi más profundo y eterno agradecimiento a mis padres, los dos pilares fundamentales de mi vida, la fuerza y el consuelo, el guía y el abrazo. A ustedes les debo no solo este logro, sino la persona que soy hoy. Me faltará vida para devolverles una mínima parte de todo el amor, el apoyo y las oportunidades que me han brindado. Este título es tan suyo como mío, porque sin ustedes, este camino jamás hubiera sido posible. También quiero agradecer a mi querida hermanita Leonor, mi Juani, gracias por tu apoyo y compañía sin condiciones.

También agradezco con todo mi cariño a mis primas Ashley, Schlomyt y Katerine, por ser como hermanas mayores, guiándome y cuidándome siempre con su amor incondicional; y a mi tía Quena, tía Pin, tío Juan, tío Sergio y Andrecito, por acompañar con tanto orgullo y apoyo constante cada uno de mis logros, haciéndome sentir siempre querida.

Finalmente, pero no menos importante a mi compañero de vida Sebastián, gracias por haberte convertido en un pilar esencial en mi vida desde el primer momento. Gracias por tu apoyo incondicional, por demostrarme tu amor cada día y por no soltarme la mano cuando las fuerzas flaqueaban. Mi Real hasta la muerte.

Todo lo que soy y lo que logro es gracias al amor que me envuelve. A mi familia, la que me dio la vida y la que he ido encontrando en el camino: ustedes son mi mayor fortaleza, mi refugio y mi motivo.

TABLA DE CONTENIDO

RESUMEN	5
ABSTRACT	6
I. INTRODUCCIÓN.....	7
1.1 ANTECEDENTES DEL ESTUDIO.....	7
1.2 PLANTEAMIENTO DEL PROBLEMA	12
1.3 PREGUNTAS DE INVESTIGACIÓN.....	14
1.3.1 Pregunta general.....	15
1.3.2 Preguntas específicas	15
1.4 OBJETIVOS DEL ESTUDIO	15
1.4.1 Objetivo general.....	15
1.4.2 Objetivos específicos	16
II. MARCO CONCEPTUAL.....	16
2.1 NATURALEZA DE LA CURIOSIDAD	16
2.1.1 Tipos de Curiosidad: Perceptual y Epistémica	17
2.1.2 El Modelo Bidimensional de Litman: Interés y Privación (I/D).....	17
2.2 TEORÍAS FUNDAMENTALES DE LA CURIOSIDAD EPISTÉMICA.....	19
2.2.1 Teoría de la Brecha de la Información.....	19
2.2.2 Teoría de las Emociones Epistémicas.....	20
2.2.3 Perspectiva del Desarrollo Infantil y Aprendizaje Activo	20
2.3 CURIOSIDAD EPISTÉMICA EN EL CONTEXTO EDUCATIVO	23
2.4 CURIOSIDAD EN EL APRENDIZAJE DE LAS MATEMÁTICAS.....	24
2.4.1 La Curiosidad Epistémica en el Razonamiento Inferencial Informal	25
III. METODOLOGÍA.....	28
3.1 ENFOQUE Y DISEÑO METODOLÓGICO	28
3.2 PARTICIPANTES	28
3.3 DISEÑO Y CARACTERÍSTICAS DEL INSTRUMENTO	30
3.4 PROCEDIMIENTO DE VALIDACIÓN	32
3.4.1 Diseño de la escala.....	32
3.4.2 Validación de Contenido.....	33
3.4.3 Revisión de jueces expertos.	34
3.4.4 Entrevistas cognitivas.	39
3.4.5 Procedimiento de análisis y revisión de ítems.	42
3.4.6 Aplicación Piloto	43
3.4.8 Programa de datos que se utilizara en el estudio	46

3.5 CONSIDERACIONES ÉTICAS	47
IV. ANÁLISIS Y RESULTADOS	48
4.1 RESULTADOS DERIVADOS DE LAS ENTREVISTAS COGNITIVAS	48
4.1.1 Comprensión semántica y lingüística	50
4.1.3 Comprensión conceptual.....	54
4.1.4 Engagement y pertinencia percibida.....	56
4.2 RESULTADOS DE VALIDACIÓN DEL CONTENIDO MEDIANTE EL JUICIO DE EXPERTOS	58
4.2.1 Evaluación inicial de la validez de contenido por jueces expertos	59
4.2.2 Claridad.....	61
4.2.3 Pertinencia.....	65
4.2.4 Relevancia.....	68
4.2.5. Suficiencia.....	70
4.2.6 Evaluación final de la validez de contenido tras la revisión por jueces expertos	72
4.3 RESULTADOS DE APLICACIÓN PILOTO.....	75
4.3.1 Análisis de los resultados de la aplicación piloto para segundo básico	75
4.3.2 Análisis de los resultados de la aplicación piloto en tercero básico	82
4.3.3 Análisis de los resultados de la aplicación piloto en cuarto básico	88
4.3.4 Análisis final de los resultados de la aplicación piloto.....	95
V. DISCUSIONES.....	97
5.1 Importancia y aporte del instrumento en el contexto nacional e internacional.....	98
5.2 Análisis de la validez de contenido: El rol de la V de Aiken y el juicio de expertos	99
5.3 Evaluación de los resultados piloto y su relación con las problemáticas planteadas ...	101
VI. CONCLUSIONES, LIMITACIONES Y PROYECCIONES DEL ESTUDIO	103
6.1 Conclusiones de objetivos y preguntas de investigación	103
VI. CONCLUSIONES, LIMITACIONES Y PROYECCIONES DEL ESTUDIO	104
6.1 Conclusiones de objetivos y preguntas de investigación	104
6.1.1 Conclusiones del objetivo general.....	104
6.1.2 Conclusiones del objetivo específico uno.....	105
6.1.3 Conclusiones del objetivo específico dos	106
6.2 Limitaciones y proyecciones del estudio.....	108
6.2.1 Limitaciones del estudio	108
6.2.2 Proyecciones del estudio	111
REFERENCIAS.....	114
ANEXOS	127
7.1 anexo 1: manual de protocolo para la aplicación de la entrevista cognitiva.....	127
7.2 anexo 2: instrumento de validación	132
7.3 anexo 3: instrumento para la evaluación de entrevista cognitiva	136

ÍNDICE DE TABLAS

Tabla 1	31
Tabla 2	33
Tabla 3.....	35
Tabla 4.....	36
Tabla 6.....	41
Tabla 7.....	49
Tabla 8.....	59
Tabla 9.....	61
Tabla 10.....	64
Tabla 11.....	65
Tabla 12.....	67
Tabla 13.....	68
Tabla 14.....	70
Tabla 15.....	70
Tabla 16.....	72
Tabla 17.....	72
Tabla 18.....	76
Tabla 19.....	77
Tabla 20.....	77
Tabla 21.....	78
Tabla 22.....	79
Tabla 23.....	82
Tabla 24.....	83
Tabla 25.....	84
Tabla 26.....	85
Tabla 27.....	86
Tabla 28.....	89
Tabla 29.....	90
Tabla 30.....	91
Tabla 31.....	92
Tabla 32.....	93

ÍNDICE DE ILUSTRACIONES

Figura 1.....	13
Figura 2.....	22
Figura 3.....	47
Figura 4.....	52
Figura 5.....	55
Figura 6.....	63
Figura 7.....	63
Figura 8.....	73

RESUMEN

Este estudio diseñó y validó una escala cuantitativa para medir la curiosidad epistémica en matemáticas en estudiantes de segundo a cuarto básico en Chile, basada en el modelo bidimensional de Litman (Tipo-I: interés; Tipo-D: privación) y contextualizada en situaciones de razonamiento inferencial informal. La validación incluyó juicio de expertos (V de Aiken entre 0.80 y 1.00) y entrevistas cognitivas, que identificaron y corrigieron problemas lingüísticos y conceptuales. Una aplicación piloto reveló variabilidad y discriminación aceptables en seis ítems, mientras que tres requirieron reformulación. El instrumento demuestra viabilidad teórica y metodológica, ofreciendo una herramienta robusta para investigación y diagnóstico educativo en matemáticas tempranas.

PALABRAS CLAVES

Curiosidad epistémica, validación de contenido, entrevistas cognitivas, educación matemática, psicometría.

ABSTRACT

This study designed and validated a quantitative scale to measure epistemic curiosity in mathematics among second to fourth grade Chilean students, based on Litman's bidimensional model (Type-I: interest; Type-D: deprivation) and contextualized in informal inferential reasoning scenarios. Validation involved expert judgment (Aiken's V between 0.80 and 1.00) and cognitive interviews, which identified and corrected linguistic and conceptual issues. A pilot application showed acceptable variability and discrimination for six items, while three required revision. The instrument demonstrates theoretical and methodological feasibility, providing a robust tool for research and educational assessment in early mathematics.

KEYWORDS

Epistemic curiosity, content validity, cognitive interviews, mathematics education, psychometric validation.

I. INTRODUCCIÓN

1.1 ANTECEDENTES DEL ESTUDIO

La curiosidad ha sido abordada desde múltiples disciplinas, tales como la psicología, la pedagogía, la neurociencia y las ciencias cognitivas. No existe una única definición consensuada, sino una diversidad de enfoques que reconocen su carácter multifacético. En términos generales, suele entenderse como una disposición a buscar información, enfrentarse a lo desconocido y explorar el entorno, motivada por el deseo de reducir la incertidumbre o llenar vacíos en el conocimiento (Berlyne, 1966; Chouinard et al., 2007; Jirout y Klahr, 2012; Langevin, 1971; Loewenstein, 1994; McNary, 2023).

Desde una perspectiva psicológica, la curiosidad se ha conceptualizado como un mecanismo adaptativo de búsqueda de información que no solo responde a estímulos externos, sino que también está guiado por motivaciones internas y afectivas (Berlyne, 1954, 1966; Litman y Spielberger, 2003; McNary, 2023; Renner, 2006). Si bien todos los seres humanos comparten este deseo de adquirir nueva información y conocimiento, existen diferencias individuales en la intensidad de ese deseo y en los comportamientos exploratorios resultantes (Loewenstein, 1994; McNary, 2023; Schmalty y Schneider, 2009; Spielberger y Starr, 1994).

De acuerdo con la Teoría de la Brecha de Información, la curiosidad se activa cuando el individuo percibe una carencia de conocimiento, generando un estado aversivo que motiva la búsqueda activa de información para restablecer el equilibrio cognitivo interno (Jirout et al., 2022; Loewenstein, 1994; McNary, 2023).

La propuesta de Litman y Jimerson (2004) sobre los tipos I y D de curiosidad activada por el interés potencial o por la privación de información consolida una visión integradora que, junto a otros enfoques teóricos, caracteriza a la curiosidad como un proceso cognitivo, emocional y motivacional orientado a la reducción de incertidumbre y al descubrimiento de nueva información. Esta naturaleza multifacética explica por qué la curiosidad se considera un motor esencial en el aprendizaje humano, especialmente durante la infancia.

Desde esta perspectiva, la curiosidad no puede entenderse como un elemento accesorio del proceso educativo, sino como un componente estructural y necesario para el aprendizaje significativo (Gopnik et al., 2015; Legare, 2014). Este enfoque se alinea con los actuales marcos curriculares que promueven una enseñanza centrada en el estudiante y en la activación de procesos cognitivos complejos.

En el contexto chileno, la necesidad de abordar variables motivacionales como la curiosidad epistémica se vuelve urgente al contrastar estos lineamientos con la realidad del rendimiento académico. Los últimos resultados de las pruebas estandarizadas revelan un estancamiento en el aprendizaje matemático. Por ejemplo, el Simce 2022 En Matemática se registró una baja considerable de 10 puntos en 4° básico (250), en el caso de 6to año se comparó el promedio Nacional de Matemáticas del año 2018 y 2024, dicho puntaje disminuyó de 251 a 245 (Agencia de Calidad de la Educación, 2023). A nivel internacional, Chile se posiciona por debajo del promedio de la OCDE en la prueba PISA

2022, con una brecha significativa entre el desempeño en matemáticas y el de países con sistemas educativos similares (OECD, 2023).

El MINEDUC (2018, 2024) ha enfatizado en sus Bases Curriculares y el Marco para la Buena Enseñanza la imperiosa necesidad de fomentar la indagación, la reflexión y la resolución de problemas. Estos principios se alinean con la evidencia internacional que identifica la curiosidad epistémica como un factor crucial para revertir las tendencias de bajo rendimiento, emergiendo no como un complemento, sino como un componente estructural del aprendizaje (Jirout, 2020). La investigación contemporánea la conceptualiza como un motor interno que podría explicar, más allá de la capacidad cognitiva, las diferencias en el engagement y el logro académico en matemáticas (Mussel, 2022; Shah et al., 2018).

El desarrollo de la curiosidad en el ámbito escolar se vincula con la creación de oportunidades para explorar, cuestionar y construir saberes de manera activa (Hidi y Renninger, 2006; Jirout, 2020; Kashdan y Steger, 2007). Esta vinculación se refleja en que los niños más curiosos presentan mejores niveles de autorregulación, mayor persistencia ante la dificultad y una comprensión conceptual más profunda. Esta relación, significativamente, se mantiene incluso al controlar factores como el esfuerzo o la capacidad cognitiva, lo que sugiere que la curiosidad actúa como un factor independiente en el rendimiento académico (Mussel, 2022; von Stumm et al., 2011).

En el ámbito de las matemáticas, el reconocimiento del papel de la curiosidad ha sido históricamente limitado (Chi et al., 1994; Jirout, 2020; Martin y Schwartz, 2004). A

pesar de que esta disciplina ofrece múltiples oportunidades para la formulación de preguntas y la resolución de problemas, la enseñanza tradicional ha privilegiado la transmisión de contenidos por sobre la exploración activa, limitando el potencial de las matemáticas como espacio de descubrimiento y pensamiento creativo (Chi et al., 1994; Jirout, 2020; Martin y Schwartz, 2004).

En los últimos años, diversos estudios han comenzado a reconceptualizar el vínculo entre curiosidad epistémica y aprendizaje matemático (Chi y Wylie, 2014; Litman et al., 2005; Tang et al., 2024). Esta nueva perspectiva revela que los estudiantes que sienten interés por resolver incertidumbres matemáticas o comprender el sentido de un procedimiento muestran mayor implicación cognitiva y comprensión profunda. Dicho proceso se ve potenciado porque las emociones epistémicas que acompañan a la curiosidad actúan como catalizadores del aprendizaje, activando procesos de inferencia, comparación y elaboración (Bosch y D’Mello, 2017; Pekrun et al., 2017).

En consonancia con los lineamientos del MINEDUC (2012), se requiere avanzar hacia prácticas pedagógicas que generen condiciones para que emerjan preguntas genuinas, se valoren los errores como oportunidades para investigar y se promueva el deseo de descubrir relaciones matemáticas. Las Bases Curriculares (2012) destacan la necesidad de despertar la curiosidad, fomentar la búsqueda de información y fortalecer la confianza en los propios procesos de aprendizaje. Así, la curiosidad puede actuar como punto de partida para el aprendizaje matemático y como eje organizador de experiencias cognitivas más significativas.

A nivel internacional investigaciones como las de Ahmad y Siew (2021) y Tang et al. (2024) han desarrollado escalas para medir la curiosidad en distintas etapas del desarrollo, aunque evidencian limitaciones psicométricas, especialmente en la medición de la curiosidad infantil, también investigadores como Sánchez et al. (2023) validaron por primera vez la escala de Litman "Epistemic Curiosity Scale" (2008) en hispanohablantes, la cual cuenta con limitaciones importantes y por último Seda et. al 2023 realizó la adaptación de una escala epistémica al contexto Turco para niños pero esta era respondida por los tutores legales.

En síntesis, los antecedentes revelados evidencian la importancia e impacto positivo que genera el desarrollo de la curiosidad epistémica para generar aprendizaje significativo (Jirout, 2020; von Stumm et al., 2011) y el vacío crítico en la medición de este constructo en el dominio específico de las matemáticas durante la primaria (2° a 4° básico) que es la etapa en donde la curiosidad se manifiesta y desarrolla de mayor forma, siendo incluso para educadores algo esencial en la preparación escolar (Heaviside et al., 1993; West et al., 1993). Los instrumentos existentes son genéricos o presentan limitaciones psicométricas para esta población (Sanchez et al., 2023), lo que impide captar las manifestaciones únicas de la curiosidad matemática, como la fascinación por los patrones o el disfrute en la resolución de problemas no algorítmicos.

1.2 PLANTEAMIENTO DEL PROBLEMA

En Abril de 2025 se realizó una revisión exploratoria sistemática en Web of Science con el siguiente código: TS=((curios* OR "epistemic curiosity" OR "trait

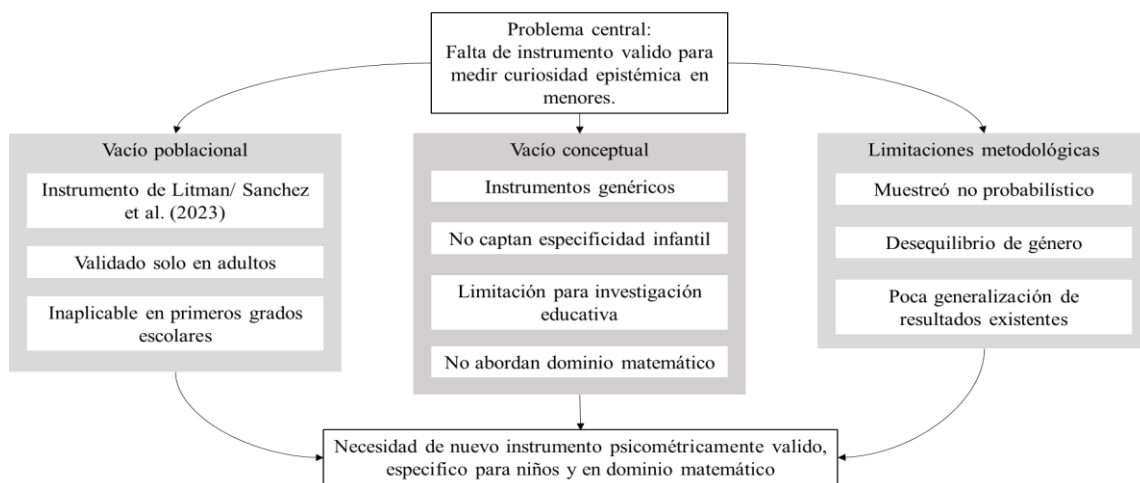
curiosity") AND (scale* OR questionnaire* OR invent* OR "self-report" OR "behavioral measure*" OR "task-based measure*" OR "physiological measure*") AND (valid* OR reliab* OR "psychometric*" OR "factor structure" OR "internal consistency"). Además, aplicando criterios rigurosos para identificar instrumentos de curiosidad en la infancia con propiedades psicométricas validadas se realizó una búsqueda en Google Académico, Scielo y Scopus, identificando únicamente 3 documentos relevantes de un total de **52** artículos detectados inicialmente. De este corpus, sólo un instrumento demostró utilidad directa para operacionalizar el constructo de curiosidad epistémica: la Escala de Curiosidad Epistémica de Litman, validada recientemente para población hispanohablante por Sánchez et al. (2023). Si bien este estudio confirma la solidez del modelo bidimensional Interés/Privación (I/D) de Litman (2008) en contexto hispano, simultáneamente revela un vacío crítico para la investigación (Figura 1). La escala fue diseñada y validada exclusivamente para población adulta, adicionalmente presenta un desequilibrio en la distribución por género en su muestreo y no utilizó técnicas de muestreo probabilístico (Bengart et al., 2018), lo que limita la generalización de sus resultados y la hace metodológica y conceptualmente inaplicable para estudiantes de los primeros grados escolares.

La revisión evidenció que algunos instrumentos existentes evalúan la curiosidad de manera genérica, sin captar las manifestaciones específicas que emergen de las actitudes de los niños hacia la curiosidad epistémica, auto curiosidad y factores relacionados con emociones o personalidad (Ahmad y Siew, 2021). Esta limitación se

declina aún más al no encontrar instrumentos que focalicen la curiosidad epistémica en el dominio matemático.

Figura 1

Dimensiones del vacío identificado en la revisión



Nota. Elaborada por los autores para esta investigación.

La carencia de un instrumento específico, válido y confiable genera tres problemas fundamentales. (1) Para la investigación, resulta imposible determinar con precisión cómo la curiosidad epistémica matemática se relaciona con variables claves como el rendimiento académico, la autorregulación del aprendizaje o la eficacia de metodologías didácticas innovadoras en el contexto local. (2) Para la práctica pedagógica, los docentes carecen de herramientas diagnósticas que les permitan identificar los perfiles de curiosidad (Tipo-I/D) de sus estudiantes, lo que impide diseñar intervenciones diferenciadas que potencien este motor esencial del aprendizaje profundo. (3) Para la política educativa, sin instrumentos contextualizados, es inviable evaluar el impacto real de las reformas curriculares que promueven el desarrollo de la curiosidad (MINEDUC,

2018, 2024) o generar evidencia local que fundamente decisiones pedagógicas basadas en datos.

Para subsanar esta brecha metodológica, este estudio se propone diseñar y validar un instrumento contextualizado que, tomando como base teórica el modelo I/D de Litman (2008), que permita medir de manera precisa y pertinente la curiosidad epistémica en escenarios de razonamiento inferencial informal en matemáticas, en estudiantes de segundo a cuarto básico en Chile.

El aporte original de esta investigación trasciende la mera adaptación lingüística. Mientras la escala original de Litman y su validación al español miden el constructo como un rasgo general en adultos, este estudio contextualiza la curiosidad epistémica en un dominio de conocimiento específico (matemáticas) y la ancla en un tipo de razonamiento particular (inferencial informal). Esta doble contextualización constituye la contribución novedosa, permitiendo capturar manifestaciones situadas y ecológicamente válidas, como el interés por descifrar patrones (Tipo-I) o la incomodidad ante una conjetura no resuelta (Tipo-D), en escenarios de exploración matemática propios del currículum nacional. En consecuencia, este instrumento no solo generará la primera evidencia empírica local sobre el rol de la curiosidad epistémica en el aprendizaje matemático inicial, sino que proveerá a educadores e investigadores de una herramienta diagnóstica específica, válida y fiable para fundamentar el diseño de intervenciones pedagógicas que potencien este motor esencial del aprendizaje profundo.

1.3 PREGUNTAS DE INVESTIGACIÓN

Tomando en cuenta lo anteriormente descrito, la presente investigación se plantea las siguientes preguntas:

1.3.1 Pregunta general

¿Cómo diseñar y validar una escala psicométrica confiable que mida curiosidad epistémica en matemáticas en estudiantes de segundo a cuarto básico del sistema escolar chileno?

1.3.2 Preguntas específicas

¿Cuáles son los componentes teóricos y operacionales que definen la curiosidad epistémica en el contexto del aprendizaje de las matemáticas en estudiantes de segundo a cuarto básico?

¿Cuál es la validez de contenido, estructural y de criterio de la escala diseñada, en función del juicio de expertos y el análisis estadístico?

¿Qué nivel de confiabilidad presenta la escala, medida a través de consistencia interna?

1.4 OBJETIVOS DEL ESTUDIO

En función de los antecedentes anteriores este estudio se propone desarrollar un instrumento de evaluación que permita medir el nivel de curiosidad epistémica en contextos matemáticos en estudiantes de segundo a cuarto básico, adaptando al contexto del currículum chileno y fundamentando modelos teóricos actuales. Para ello, nos proponemos los siguientes objetivos:

1.4.1 Objetivo general

Diseñar y validar una escala cuantitativa, válida y confiable, para medir la curiosidad epistémica de tipo I/D en contextos de razonamiento inferencial informal en matemática, en estudiantes que cursan desde segundo a cuarto básico en Chile.

1.4.2 Objetivos específicos

Identificar instrumentos existentes para medir la curiosidad epistémica en contextos matemáticos.

Construir una escala cuantitativa para medir la curiosidad epistémica, en matemática, en estudiantes que cursan segundo a cuarto básico en Chile.

Validar la escala cuantitativa para medir la curiosidad epistémica, en matemática, en estudiantes que cursan de segundo a cuarto básico en Chile.

II. MARCO CONCEPTUAL

2.1 NATURALEZA DE LA CURIOSIDAD

La curiosidad ha sido conceptualizada como un impulso cognitivo y emocional fundamental que motiva a los individuos a buscar información novedosa, explorar lo desconocido y resolver estados de incertidumbre (Jirout y Klahr, 2012; Loewenstein, 1994). Desde una perspectiva psicológica, se la considera un mecanismo adaptativo de búsqueda de información que no solo responde a estímulos externos, sino que está guiado por motivaciones internas y afectivas (Berlyne, 1954; Litman y Spielberger, 2003). Si bien todos los seres humanos comparten esta disposición a conocer, existen notorias diferencias individuales en su intensidad y en las conductas exploratorias resultantes, lo que sugiere que la curiosidad también puede entenderse como un rasgo de personalidad (Loewenstein, 1994; Spielberger y Starr, 1994).

La naturaleza de la curiosidad es intrínsecamente multifacética, Berlyne (1954) postuló que no se trata de una entidad unitaria, sino de una categoría que engloba diversas manifestaciones y formas, cada una con sus propios desencadenantes y expresiones conductuales. Esta complejidad explica por qué la curiosidad ha sido abordada desde múltiples disciplinas como la psicología, la pedagogía, la neurociencia y las ciencias cognitivas, ofreciendo una diversidad de enfoques que, en su conjunto, enriquecen la comprensión de este constructo (Berlyne, 1966; Chouinard et al., 2007; Jirout y Klahr, 2012; Langevin, 1971; Loewenstein, 1994; McNary, 2023).

2.1.1 Tipos de Curiosidad: Perceptual y Epistémica

La distinción clásica de Berlyne (1954) entre curiosidad perceptual y epistémica resulta fundamental para delimitar el objeto de este estudio. La curiosidad perceptual es provocada por estímulos sensoriales novedosos, complejos o ambiguos, y conduce a una conducta exploratoria

sensorial. Este tipo de curiosidad, observable tanto en animales como en humanos, tiene como impulso central el "experimentar y sentir", y se satisface mediante la manipulación y exploración directa del entorno (Berlyne, 1966).

En contraste, la curiosidad epistémica (CE) se orienta de manera específica hacia la adquisición de conocimiento abstracto y conceptual. Berlyne (1954) la describió como un "impulso de saber" exclusivamente humano (p. 187) que motiva la indagación y la experimentación reflexiva, subyaciendo al desarrollo intelectual y al logro académico (Berlyne, 1966, 1971). Mientras la curiosidad perceptual se centra en lo sensorial, la epistémica se interesa por las ideas, las explicaciones y la resolución de incertidumbres conceptuales. Es esta última, por su directa vinculación con los procesos de aprendizaje formal y la construcción de conocimiento, la que constituye el foco central de la presente investigación.

2.1.2 El Modelo Bidimensional de Litman: Interés y Privación (I/D)

El modelo de Litman y sus colaboradores representa un avance crucial en la operacionalización de la curiosidad epistémica. Basándose en el trabajo de Berlyne (1954), Litman y Spielberger (2003), sentaron las bases al conceptualizar la CE no sólo como un estado transitorio, sino como un rasgo de personalidad estable. Posteriormente, Litman y Jimerson (2004) profundizaron en este modelo mediante análisis psicométricos rigurosos, lo que culminó en el desarrollo de la Teoría I/D de la Curiosidad. Esta línea de investigación ha sido continuada y validada en contextos aplicados por autores contemporáneos como Tang et al. (2024), Sanchez et al. (2023), Sarac et al. (2023), Piotrowski et al. (2014) asociando consistentemente el constructo a estados de interés emocional positivo y al placer intrínseco de aprender.

Al explorar en profundidad sus componentes, Litman y Jimerson (2004) desarrollaron la Teoría I/D de la Curiosidad, concluyendo mediante análisis psicométricos rigurosos que la CE está compuesta por dos dimensiones motivacionales distintas, aunque potencialmente correlacionadas (Litman, 2005, 2008):

Curiosidad Epistémica Tipo-I (Interés): Se refiere a la curiosidad motivada por el placer anticipado de descubrir, explorar y asimilar nuevas ideas o información. Está asociada con una afectividad positiva, un interés intrínseco por el aprendizaje y la búsqueda de estimulación intelectual (Litman, 2008; Sánchez et al., 2023). Se activa predominantemente ante la novedad conceptual, disfruta del proceso de descubrimiento en sí mismo y busca nueva información primordialmente para generar y mantener afectos positivos. Litman (2008) la asocia con un objetivo de aprendizaje orientado al dominio, donde la meta es alcanzar una comprensión profunda por el simple placer y la satisfacción personal que ello reporta.

Curiosidad Epistémica Tipo-D (Privación): Hace referencia a la curiosidad motivada por la reducción de un estado afectivo negativo de incertidumbre o incomodidad cognitiva, generado por la percepción aguda de una carencia o vacío informativo específico (Litman, 2008). Se estimula cuando el individuo siente que carece de una información concreta que considera necesaria, generando un estado de tensión, ansiedad o insatisfacción. A diferencia del Tipo-I, su búsqueda está dirigida a reducir activamente esa incertidumbre puntual y restaurar la coherencia cognitiva, relacionándose con metas de aprendizaje orientadas al rendimiento, tanto de aproximación al éxito como de evitación del fracaso (Sánchez et al., 2023).

Esta investigación se basará en Litman (2007) y enmarca estas dos dimensiones en marcos teóricos más amplios: la curiosidad como estado de excitación óptima (Tipo-I) y la curiosidad

como impulso de reducción de la privación (Tipo-D). Este modelo bidimensional, que captura de forma precisa las fuentes motivacionales duales del "placer por aprender" y el "alivio por comprender", constituye el marco estructural fundamental para el diseño del instrumento de esta investigación.

2.2 TEORÍAS FUNDAMENTALES DE LA CURIOSIDAD EPISTÉMICA

2.2.1 Teoría de la Brecha de la Información

Postulada por Loewenstein (1994), esta teoría explica el mecanismo de activación cognitiva de la curiosidad. Sostiene que esta se desencadena cuando un individuo percibe una discrepancia o "brecha" entre lo que sabe y lo que desea o necesita saber, generando un estado de arousal aversivo que motiva la búsqueda específica de información para restaurar el equilibrio cognitivo. Esta teoría provee el fundamento directo para la dimensión de Privación (Tipo-D) del modelo de Litman, conceptualizando la curiosidad como un estado similar a una necesidad insatisfecha que impulsa una exploración focalizada para "cerrar la brecha" y reducir la tensión (Litman, 2008). Características como su intensidad, transitoriedad e impulsividad (Loewenstein, 1994) explican la urgencia y el carácter dirigido de este tipo de curiosidad. Desde una perspectiva neurocientífica, esta búsqueda activa se ha asociado con la activación de circuitos de recompensa, lo que explica el componente placentero de la saciedad cognitiva (Liu, 2024; Oudeyer, 2013).

2.2.2 Teoría de las Emociones Epistémicas

Desde este marco, prominentemente desarrollado por Pekrun y sus colegas (2017, 2018), la curiosidad es reconocida como una emoción epistémica central. Estas emociones surgen específicamente durante actividades de generación y procesamiento de conocimiento, tales como

la lectura, la resolución de problemas o la reflexión, y están directamente ligadas a la calidad y el resultado de estos procesos cognitivos. La curiosidad se activa típicamente ante lagunas informativas, incongruencias o cuando la nueva información desafía las concepciones previas del estudiante (Pekrun et al., 2018). Suele presentarse al inicio de una secuencia emocional epistémica: la información novedosa o contradictoria puede generar sorpresa o confusión, y si esta última se mantiene, puede derivar en ansiedad o frustración. Sin embargo, si la incógnita se resuelve satisfactoriamente, el estudiante experimenta disfrute y una curiosidad renovada (Bosch y D'Mello, 2017; Pekrun et al., 2018). Esta teoría enfatiza la naturaleza profundamente afectivo-cognitiva de la curiosidad y su papel catalizador en procesos superiores como la inferencia, la comparación y la elaboración mental.

2.2.3 Perspectiva del Desarrollo Infantil y Aprendizaje Activo

Desde la psicología del desarrollo, la curiosidad es considerada un motor central del aprendizaje infantil. Los niños son exploradores naturales, impulsados por una tendencia intrínseca a hacer preguntas, experimentar, observar minuciosamente y buscar explicaciones causales para los fenómenos que no comprenden (Chouinard et al., 2007; Gopnik et al., 2015). En este sentido, la curiosidad no es un elemento accesorio o complementario al aprendizaje, sino un componente estructural que orienta la atención, organiza el pensamiento y promueve la construcción activa de significados (Legare, 2014). La conexión entre curiosidad y aprendizaje ha sido también abordada desde el punto de vista de la motivación y la cognición. Investigaciones demuestran que una mayor curiosidad, especialmente cuando se relaciona con niveles manejables de incertidumbre, se asocia con una mejor codificación de la información en la memoria y una mayor capacidad de

transferencia de conocimientos a nuevas situaciones, lo cual es un indicador clave de aprendizaje significativo y profundo (Gopnik et al., 2015; Jirout, 2020).

En conjunto, estas teorías proporcionan una base integral y complementaria para entender la curiosidad epistémica, cuya integración puede visualizarse en el siguiente modelo conceptual (Figura 2). La Teoría de la Brecha de Información (Loewenstein, 1994) explica el mecanismo de activación cognitiva fundamental: la percepción de una discrepancia entre el conocimiento actual y el deseado. El modelo I/D de Litman (2008) opera como el marco integrador central que captura de manera operativa las dos fuentes motivacionales duales que surgen de esta brecha: la búsqueda de placer (Tipo-I) y la búsqueda de alivio (Tipo-D).

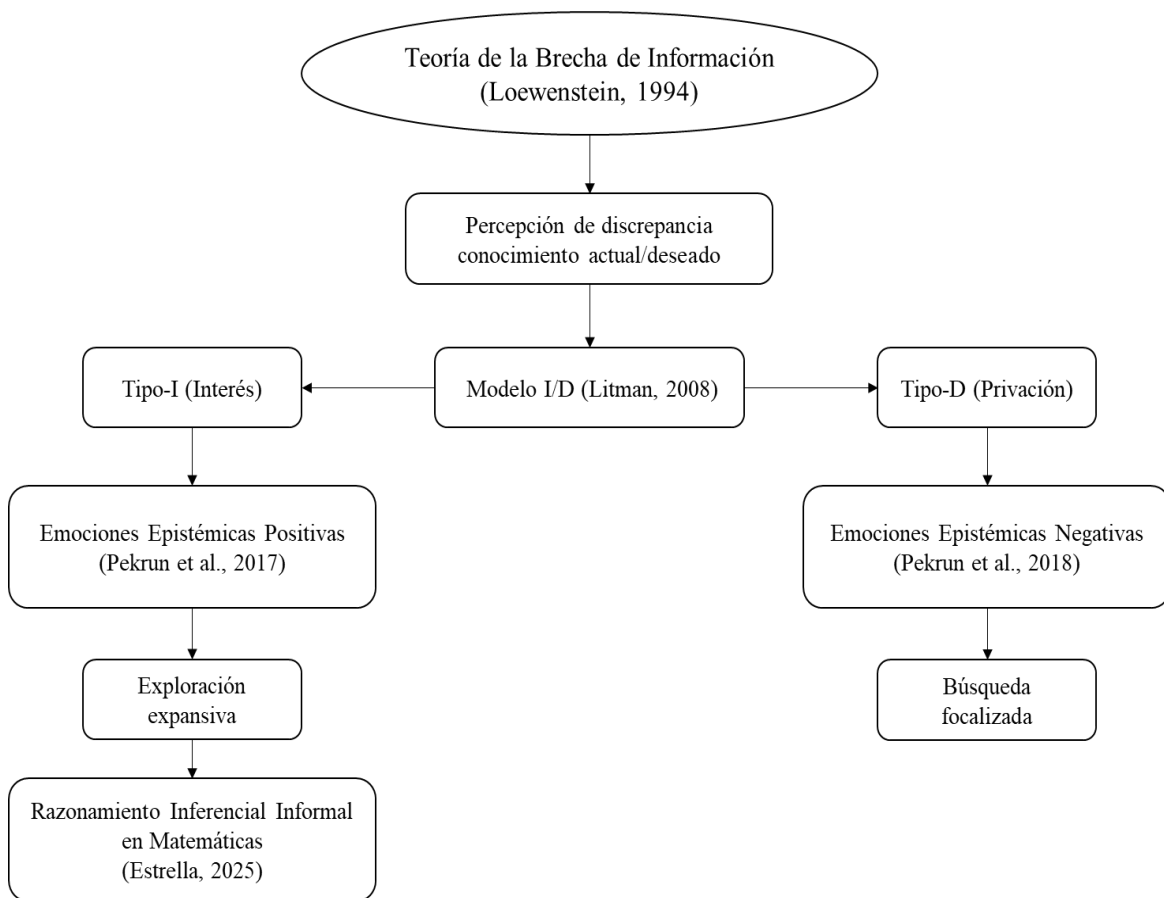
Es precisamente en esta dicotomía motivacional donde la Teoría de las Emociones Epistémicas (Pekrun et al., 2017, 2018) encuentra su mayor resonancia, describiendo la naturaleza afectiva de la curiosidad y su rol dinámico en el procesamiento cognitivo. La integración entre ambos modelos revela una complementariedad fundamental para comprender la curiosidad en contextos matemáticos: la Curiosidad Tipo-I (Interés) se asocia con emociones epistémicas positivas como el disfrute intelectual y la fascinación, las cuales emergen cuando los estudiantes experimentan el placer de descubrir relaciones matemáticas por sí mismos. Por el contrario, la Curiosidad Tipo-D (Privación) se vincula con emociones epistémicas de valencia negativa como la confusión o la perplejidad, que surgen ante lagunas informativas específicas en la comprensión de conceptos matemáticos (Pekrun et al., 2017).

Finalmente, la Perspectiva del Desarrollo Infantil (Gopnik et al., 2015; Legare, 2014) sitúa este impulso integrado (cognitivo-motivacional-afectivo) como un motor fundamental y constitutivo del aprendizaje activo durante la infancia. En el contexto del razonamiento inferencial

informal en matemáticas, esta dinámica se activa cuando los estudiantes se enfrentan a problemas que requieren generar conjeturas, identificar patrones o establecer relaciones no inmediatamente evidentes, constituyendo un motor esencial para la persistencia en la resolución de problemas matemáticos complejos (Jirout, 2020; Martin y Schwartz, 2004).

Figura 2

Modelo Integrado de Curiosidad Epistémica en Razonamiento Inferencial Matemático.



Nota. Elaborada por los autores para esta investigación.

2.3 CURIOSIDAD EPISTÉMICA EN EL CONTEXTO EDUCATIVO

Diversos autores coinciden en que el aprendizaje que se deriva de la curiosidad es cualitativamente superior a la mera recepción pasiva de contenidos. Según Chi y Wylie (2014), el razonamiento metacognitivo impulsado por la curiosidad apoya la activación de conocimientos previos, la elaboración de inferencias y la reorganización de las representaciones mentales, procesos todos ellos asociados a una comprensión profunda. Esta idea es coherente con los principios del aprendizaje activo, que postulan que la participación cognitiva significativa y el procesamiento constructivo conducen a aprendizajes más duraderos y transferibles (Bonwell y Eison, 1991).

Empíricamente, la curiosidad se asocia positivamente con comportamientos y actitudes que favorecen el aprendizaje, como un mayor compromiso o "engagement", una persistencia incrementada ante los desafíos, una mayor disposición para explorar nuevas ideas y una capacidad más desarrollada para autorregular el propio aprendizaje (Hidi y Renninger, 2006; Jirout et al., 2022; Kashdan y Steger, 2007). Estos factores han sido vinculados de forma consistente con un mejor rendimiento académico sostenido a lo largo del tiempo. Investigaciones longitudinales y metaanálisis sugieren que la curiosidad actúa incluso como un predictor significativo del éxito académico en etapas tempranas, manteniendo su poder predictivo al controlar variables como el esfuerzo o la capacidad cognitiva tradicional, lo que la erige como un factor explicativo independiente y de gran relevancia educativa (Shah et al., 2018; von Stumm et al., 2011; Mussel, 2022).

2.4 CURIOSIDAD EN EL APRENDIZAJE DE LAS MATEMÁTICAS

En el ámbito específico de las matemáticas, existe un consenso creciente sobre la necesidad de resignificar su enseñanza, transitando desde una lógica de transmisión y aplicación de

procedimientos hacia una lógica de indagación guiada, exploración conceptual y búsqueda de sentido (Jirout, 2020; Martin y Schwartz, 2004). Esta transformación metodológica es crucial para alinear la enseñanza matemática con los principios del aprendizaje activo, en los que la curiosidad del estudiante es reconocida y estimulada como una fuente legítima y poderosa de motivación para aprender.

La evidencia empírica reciente sugiere que la participación activa del estudiante en actividades matemáticas con un enfoque exploratorio puede potenciar específicamente su curiosidad epistémica. Esta forma de curiosidad se manifiesta de manera palpable cuando el alumno no se conforma con aplicar una fórmula de manera mecánica, sino que intenta resolver una incertidumbre específica, comprender el porqué subyacente a un procedimiento o descubrir las relaciones estructurales que gobiernan un problema (Tang et al., 2024). En este contexto, la curiosidad puede surgir tanto de la necesidad de resolver un problema que no se comprende del todo (acercándose al Tipo-D) como del placer estético e intelectual de descubrir nuevas relaciones, patrones o demostraciones elegantes (acercándose al Tipo-I), respondiendo así a las dos formas de curiosidad epistémica descritas en la literatura (Litman et al., 2005).

2.4.1 La Curiosidad Epistémica en el Razonamiento Inferencial Informal

El razonamiento inferencial informal constituye un dominio particularmente fértil para la observación y medición de la curiosidad epistémica en matemáticas. Este tipo de razonamiento se refiere a los procesos mediante los cuales los estudiantes, sin apelar inicialmente a reglas formales o algoritmos establecidos, elaboran conjeturas fundamentadas, predicen resultados, generalizan a partir de la observación de casos particulares y buscan patrones para construir sentido y coherencia en situaciones matemáticas novedosas (Stylianides y Stylianides, 2017; Ellis, 2007).

La naturaleza inherentemente exploratoria, conjetural y abierta de este razonamiento lo convierte en un escenario ideal para la activación de la curiosidad epistémica. Cuando un estudiante se enfrenta a una situación que demanda una inferencia como descifrar la regla subyacente en una secuencia, predecir la continuidad de un patrón geométrico o conjeturar sobre la relación entre variables, la "brecha de información" (Loewenstein, 1994) se hace palpable. Es en este espacio de incertidumbre conceptual donde la curiosidad epistémica emerge como un motor esencial para la indagación (Jirout y Klahr, 2012).

En este contexto, las dos dimensiones de la curiosidad epistémica propuestas por Litman (2008) encuentran manifestaciones distintivas:

- La Curiosidad Tipo-I (Interés por el conocimiento) se manifiesta como el disfrute intrínseco de explorar posibilidades, la fascinación ante la emergencia de un patrón y el placer lúdico de manipular ideas matemáticas
- La Curiosidad Tipo-D (Privación de conocimiento) se activa ante la incomodidad cognitiva generada por una incógnita, la tensión motivacional por verificar una conjetura o la necesidad imperiosa de resolver una ambigüedad para restaurar la coherencia lógica.

Aplicado al nivel de primaria, este marco permite operacionalizar ambas dimensiones. Para estudiantes de este ciclo, la Curiosidad Tipo-I (Interés) se puede observar en: (a) el disfrute intrínseco al descubrir patrones numéricos o geométricos; (b) la fascinación por explorar múltiples soluciones o caminos; (c) la valoración positiva al encontrar relaciones inesperadas; y (d) la motivación para indagar de forma autónoma más allá de lo solicitado. Por su parte, la Curiosidad Tipo-D (Privación) se expresa mediante: (a) la incomodidad ante una conjetura no resuelta; (b) la persistencia focalizada en cerrar una brecha de comprensión específica; (c) la búsqueda

insistente de información para eliminar una incertidumbre; y (d) la necesidad de resolver contradicciones aparentes o lagunas en su razonamiento (adaptado de Litman, 2008).

Esta articulación teórica se concreta en el currículum nacional. En el contexto chileno, el razonamiento inferencial informal se promueve en actividades como: predecir términos en secuencias numéricas (OA 9, 2° básico), formular conjeturas sobre propiedades de figuras 2D y 3D (OA 13, 3° básico), o inferir relaciones en problemas de datos (OA 19, 4° básico) (Ministerio de Educación de Chile, 2018).

En síntesis, la integración de este marco conceptual permite comprender la curiosidad epistémica no como un mero antecedente, sino como un proceso constitutivo del razonamiento inferencial informal en matemáticas. Su manifestación bidimensional (Litman, 2008) provee una lente para analizar cómo los estudiantes se involucran cognitivamente y afectivamente con problemas que requieren conjeturar, generalizar y buscar patrones. Es sobre esta base teórica sólidamente articulada que vincula la psicología de la curiosidad (Loewenstein, 1994; Litman, 2008) con la didáctica del razonamiento matemático (Stylianides y Stylianides, 2017; Ellis, 2007) que se diseñará un instrumento de medición específico y contextualizado para estudiantes de educación primaria.

A partir de la base conceptual expuesta, se diseñará un instrumento específico que permita medir la curiosidad epistémica en contextos matemáticos, considerando el desarrollo cognitivo y emocional de los estudiantes de educación primaria.

El capítulo siguiente presenta el diseño metodológico que orientará la creación y validación de dicha escala, detallando las fases y procedimientos que guiarán su desarrollo.

III. METODOLOGÍA

3.1 ENFOQUE Y DISEÑO METODOLÓGICO

La presente investigación adopta un enfoque cuantitativo, de tipo instrumental, cuyo propósito es diseñar y validar una escala para medir la curiosidad epistémica en estudiantes de segundo a cuarto básico en el área de matemáticas específicamente en el razonamiento inferencial informal. Se emplea un diseño no experimental, transversal y descriptivo, dado que no se manipulan variables independientes, sino que se observa el comportamiento psicométrico de un instrumento construido a partir de fundamentos teóricos y antecedentes empíricos previos (Aracena et al., 2024 y Jirout y Klahr, 2012).

Este tipo de estudio responde a la necesidad de desarrollar herramientas válidas y confiables para la medición de constructos psicológicos complejos en contextos escolares, siguiendo recomendaciones de autores como McNary (2023), Tang et al. (2024) Litman y Jimerson (2004) quienes señalan que la precisión conceptual y psicométrica en la evaluación de la curiosidad epistémica es un paso esencial para su comprensión y promoción pedagógica.

3.2 PARTICIPANTES

En este estudio participaron 100 estudiantes de segundo a cuarto año básico, pertenecientes a establecimientos urbanos municipales de la comuna de Los Ángeles, de la provincia del Biobío, Chile. La selección de la muestra se realizará mediante un muestreo no probabilístico intencionado (Ahmad y Siew, 2021). Los criterios de inclusión consideran a estudiantes que cursen desde segundo a cuarto año de educación básica, limitándose geográficamente la muestra a instituciones

ubicadas en la comuna de Los Ángeles de la provincia del Biobío. Se excluyen de este estudio aquellos establecimientos que no cumplan con estas condiciones.

La elección de esta población (edades aproximadas 7-10 años) se fundamenta en una convergencia de hitos evolutivos críticos que configuran una ventana sensible para el desarrollo de la curiosidad epistémica en el dominio matemático. Desde la teoría del desarrollo cognitivo, este período corresponde a la etapa de operaciones concretas, caracterizada por la emergencia del pensamiento lógico, la capacidad de clasificación seriada y el razonamiento transitivo, todas ellas habilidades precursoras esenciales para el razonamiento inferencial informal en matemáticas (Inhelder y Piaget, 1958). Complementariamente, la evidencia neurocientífica indica que en esta etapa se acelera la mielinización de los lóbulos frontales, lo que sustenta mayores capacidades de autorregulación y persistencia ante desafíos cognitivos, componentes cruciales para sostener los estados de curiosidad epistémica durante la resolución de problemas (Giedd et al., 2012).

Específicamente en el desarrollo del razonamiento matemático, este ciclo representa un período de transición desde un pensamiento basado en procedimientos concretos hacia formas más abstractas de razonamiento inferencial. Los niños entre los 7 y 9 años desarrollan la capacidad de realizar inferencias sobre relaciones cuantitativas, identificar patrones emergentes y formular conjeturas simples, que son los componentes fundamentales del razonamiento inferencial informal (Braithwaite y Siegler, 2017). Esta transición se ve reflejada en el currículum nacional, donde en estos niveles se consolidan las habilidades de razonamiento y resolución de problemas previo a la introducción de herramientas tecnológicas que podrían mediatizar el proceso de exploración conceptual (Bases Curriculares, 2018). Finalmente, desde la perspectiva del desarrollo motivacional, esta etapa constituye una ventana crítica para la consolidación de disposiciones

epistémicas estables, donde la curiosidad comienza a manifestarse como un rasgo diferenciado y dominio-específico, alcanzando una manifestación potente y medible en contextos de exploración matemática (Gopnik et al., 2015; Jirout y Klahr, 2012).

3.3 DISEÑO Y CARACTERÍSTICAS DEL INSTRUMENTO

La escala será construida a partir de la estructura teórica de curiosidad epistémica de Litman (2008) centrada en la curiosidad epistémica basada en sus dos dimensiones: tipo I (interés) y tipo D (privación), donde utiliza subdimensiones las cuales son, (1) Motivación intrínseca para obtener información. (2) Se activa al conocer algo nuevo. (3) Disfruta de nuevos descubrimientos. (4) Buscar nueva información para estimular el afecto positivo y de Privación Tipo-D (1) Placer intrínseco y preocupaciones extrínsecas. (2) Se estimula cuando se carece de información. (3) Estado de ansiedad o insatisfacción. (4) Busca reducir estados indeseables de incertidumbre e incógnitas específicas.

Con base en el modelo teórico de Litman (2008), se diseñó una tabla de especificaciones que operacionaliza las dimensiones de curiosidad epistémica en el contexto de razonamiento inferencial informal en matemáticas. La escala contempla un total de 8 ítems distribuidos equitativamente entre las dos dimensiones principales, con 4 ítems por cada subdimensión, tal como se aprecia en la Tabla 1.

Tabla 1

Distribución de ítems por dimensión y subdimensión de la escala de curiosidad epistémica en matemáticas

Dimensión principal	Subdimensión	Definición Operacional	Nº de ítems	Ejemplo de Comportamiento Observable
Tipo-I (Interés)	Motivación intrínseca	Disposición a buscar información por el placer inherente de aprender	1	Disfruta explorando diferentes formas de resolver un problema
	Activación por novedad	Respuesta de interés ante situaciones matemáticas novedosas	1	Se entusiasma al descubrir un patrón matemático inesperado
	Disfrute del descubrimiento	Placer derivado del proceso de resolver incógnitas matemáticas	1	Siente satisfacción al comprender por qué un procedimiento funciona
	Búsqueda de afecto positivo	Exploración dirigida a mantener estados emocionales positivos	1	Persiste en la resolución porque le divierte el desafío
Tipo-D (Privación)	Preocupación por lagunas	Malestar ante la percepción de carencias informativas específicas	1	Se siente incómodo cuando no entiende un concepto matemático
	Activación por carencia	Respuesta motivacional ante la falta de información necesaria	1	Se siente impulsado a buscar información que falta para resolver el problema
	Estado de tensión cognitiva	Experiencia de incomodidad ante la incertidumbre no resuelta	1	Experimenta frustración cuando no puede verificar una conjetura
	Reducción de incertidumbre	Búsqueda dirigida a eliminar estados de ambigüedad cognitiva	1	Trabaja persistentemente hasta disipar todas sus dudas

Nota. La tabla se basa en el modelo de Litman (2008) y su adaptación al contexto de razonamiento inferencial informal en matemáticas (Estrella, 2025).

3.4 PROCEDIMIENTO DE VALIDACIÓN

3.4.1 Diseño de la escala

Para la propuesta preliminar se elaborará un ítem de respuesta cerrada con cuatro alternativas para cada una de las subdimensiones propuestas por Litman (2008). La construcción inicial de estos ítems será realizada por un equipo de tres investigadores, conformado por estudiantes de pedagogía de décimo semestre de la carrera de Educación General Básica y un doctor en didáctica de las matemáticas. Como apoyo al proceso de refinamiento, se utilizará el asistente de IA DeepSeek para analizar los borradores de ítems generados por el equipo, con el

objetivo exclusivo de obtener sugerencias sobre claridad semántica, redacción y adecuación lingüística para la población objetivo, mas no para la generación inicial del contenido, la cual es responsabilidad de los investigadores. El diseño de los escenarios se orientará a través de una propuesta de enseñanza para el razonamiento inferencial informal (Estrella, 2025).

Para garantizar que los ítems representen situaciones auténticas y ecológicamente válidas, cada uno se contextualizará en escenarios de razonamiento inferencial informal alineados explícitamente con los Objetivos de Aprendizaje (OA) del currículum nacional chileno para los niveles de 2° a 4° básico (Ministerio de Educación de Chile, 2018). Esta alineación asegura que los estímulos presentados a los estudiantes sean relevantes para su experiencia matemática escolar y permita una medición precisa de la curiosidad epistémica en su contexto natural de desarrollo. La Tabla 2 ejemplifica esta articulación entre las dimensiones teóricas, el currículum y la manifestación en ítems.

Tabla 2

Ejemplos de alineación entre dimensiones de la curiosidad epistémica y objetivos de aprendizaje del currículum chileno

Dimensión CE	Objetivo de Aprendizaje (OA)	Grado	Escenario de Ítem Ejemplo
Tipo-I	OA 9: Descubrir patrones en secuencias numéricas	2° básico	"Cuando ves una fila de números que va 2, 4, 6, 8... ¿te gusta descubrir qué número sigue y por qué?"
Tipo-D	OA 13: Formular conjeturas sobre propiedades de figuras 3D	3° básico	"Si no puedes averiguar por qué una figura rueda y otra no, ¿te molesta hasta encontrar la respuesta?"
Tipo-I	OA 19: Inferir relaciones entre variables en gráficos	4° básico	"Al ver un gráfico que muestra cómo crecen las plantas, ¿te entusiasma tratar de predecir qué pasará después?"

Nota. La tabla presenta ejemplos ilustrativos de la articulación entre el marco teórico y el currículum. Los ítems definitivos serán desarrollados siguiendo este principio de contextualización para todas las subdimensiones.

3.4.2 Validación de Contenido.

Para garantizar la validez de contenido, la escala será sometida a una revisión sistemática por un panel de seis jueces expertos (Rodríguez, 2016), especialistas en Educación Matemática y evaluación. Este proceso se complementará con entrevistas cognitivas realizadas a una submuestra de diez estudiantes de segundo a cuarto básico, donde participaran 5 mujeres y 5 hombres. El enfoque mixto permitirá triangular evidencias cualitativas y cuantitativas para asegurar que los ítems representaran fielmente el constructo teórico y fueran comprensibles para la población objetivo (Aracena et al., 2024).

La evaluación de la representatividad del constructo se articula a través de una matriz de alineación que operacionaliza tres dimensiones críticas, en coherencia con los estándares que enfatizan la necesidad de evidencias múltiples para sustentar las interpretaciones de un test (American Educational Research Association et al., 2014; Boateng et al., 2018). Primero, la representatividad teórica exige que cada ítem demuestre una conexión explícita con las dimensiones del modelo I/D de Litman (2008), especificando si mide curiosidad por interés (Tipo-I) o por privación (Tipo-D). Segundo, la representatividad contextual asegura que todos los ítems se sitúen en escenarios auténticos de razonamiento inferencial informal en estadística, alineados con los Objetivos de Aprendizaje del currículum nacional (Ministerio de Educación de Chile, 2018). Tercero, la representatividad evolutiva garantiza que la formulación de los ítems sea apropiada para el desarrollo cognitivo y lingüístico de niños entre 7 y 10 años, considerando la

transición hacia las operaciones concretas (Inhelder y Piaget, 1958). Esta aproximación multidimensional asegura que el instrumento capture las manifestaciones observables de la curiosidad epistémica dentro de su dominio específico de aplicación.

3.4.3 Revisión de jueces expertos.

El panel de expertos estará conformado por seis especialistas en Educación Matemática y evaluación, seleccionados mediante un muestreo intencional que garantizará la idoneidad para validar el instrumento. Los criterios de inclusión considerarán: (a) formación de posgrado en las áreas de especialización; (b) experiencia profesional en educación matemática o evaluación educativa; (c) conocimiento del constructo de curiosidad epistémica o variables motivacionales en el aprendizaje; y (d) experiencia en procesos de validación de instrumentos o diseño de evaluaciones educativas. La selección se realizará mediante identificación de perfiles en bases de datos académicas y referencias institucionales, seguida de una invitación formal que especificará los objetivos del estudio y los criterios de evaluación que deben considerar para realizar la evaluación correspondiente.

Una vez finalizado el diseño preliminar del instrumento para medir la curiosidad epistémica, este se enviará al panel de jueces expertos. El material entregado incluirá una tabla explicativa con las definiciones conceptuales, las dimensiones teóricas y los criterios de evaluación que deben considerar los jueces explicitadas en Tabla 3 y una rúbrica detallada tal como se aprecia en la Tabla 4 que orientará el proceso de valoración de cada ítem según los 4 criterios que se desean evaluar.

Tabla 3

Dimensiones y características del instrumento para los jueces expertos.

DIMENSIONES	CARACTERÍSTICAS	CANTIDAD DE ÍTEM	CARACTERÍSTICAS GENERALES
A. Curiosidad de Interés (I)	A1. Motivación intrínseca para obtener información	1	Curiosidad de Interés (I) Relacionadas con el deseo de adquirir conocimiento por interés genuino, disfrute y motivación intrínseca (Litman, 2008). Es el deseo de adquirir nuevo conocimiento motivado por la anticipación de la recompensa emocional positiva que genera el proceso de aprendizaje en sí mismo, induciendo sentimientos de interés y disfrute (Litman, 2005; Litman & Jimerson, 2004).
	A2. Se activa al conocer algo nuevo	1	
	A3. Disfruta de nuevos descubrimientos	1	
	A4. Buscar nueva información para estimular sentimientos positivos	1	
B. Curiosidad por Privación (D)	B1. Placer intrínseco y preocupaciones extrínsecas	1	Curiosidad por Privación (D) Relacionadas con la búsqueda de conocimiento para reducir estados de incertidumbre o ansiedad por falta de información (Litman, 2019). Se manifiesta cuando los individuos sienten que carecen de información necesaria; este déficit de conocimiento percibido produce el sentimiento de 'privación' asociado con la curiosidad. (Litman 2005). Enfocada en el desempeño o rendimiento , para las que la corrección, precisión y relevancia del conocimiento recién adquirido son aspectos fundamentales (Litman, 2008).
	B2. Se estimula cuando se carece de información.	1	
	B3. Estado de ansiedad o insatisfacción.	1	
	B4. Busca reducir estados indeseables de incertidumbre e incógnitas específicas.	1	

Nota. La tabla se basa en el modelo bidimensional de curiosidad epistémica de Litman (2008) y su adaptación al contexto de razonamiento inferencial informal en matemáticas.

Tabla 4

Descripción de categorías y sus indicadores según la calificación de los jueces expertos.

CATEGORÍA	CALIFICACIÓN	INDICADOR
Claridad El ítem es claramente comprensible lingüísticamente.	1	El ítem no es comprensible o es altamente ambiguo . Su formulación impide una evaluación consistente y requiere una reescritura completa.
	2	El ítem presenta ambigüedades significativas en su redacción y requiere bastantes modificaciones en su estructura o terminología para ser preciso y evaluable.
	3	El ítem es comprensible en su mayoría , pero algún término o frase es vago o poco específico. Necesita una modificación puntual para lograr una redacción óptima.
	4	El ítem está formulado con precisión y sin ambigüedades . La conducta o actitud que se evalúa es clara y directamente observable para el evaluador.
Pertinencia El ítem planteado es pertinente en relación con la característica de la curiosidad que está midiendo.	1	El ítem planteado no es pertinente en relación con la característica de la curiosidad que mide.
	2	El ítem planteado es parcialmente pertinente en relación con la característica de la curiosidad que mide.
	3	El ítem planteado es pertinente en relación con la característica de la curiosidad que mide.
	4	El ítem planteado es totalmente pertinente en relación con la característica de la curiosidad que mide.
Relevancia El ítem planteado es relevante de acuerdo con el objetivo de la investigación, es decir, debe ser incluida.	1	El ítem planteado no es relevante y se puede eliminar .
	2	El ítem planteado tiene cierta relevancia , pero otro ítem puede estar midiendo algo similar.
	3	El ítem planteado es relativamente relevante .
	4	El ítem planteado es muy relevante y debe ser incluido .
Suficiencia Los ítem son suficientes en relación con la dimensión que se está midiendo.	1	Los ítems planteados no son suficientes.
	2	Los ítems planteados miden algunos aspectos de la dimensión.
	3	Los ítems planteados deben aumentar uno o dos más para medir la dimensión completamente.
	4	Los ítems planteados son suficientes.

Nota. Elaborado por investigadores en base a Aracena et al., 2024.

La rúbrica fue desarrollada para estandarizar el proceso de evaluación de los ítems por parte de los jueces expertos.

Cada juez analizará los ítems de manera independiente considerando cuatro categorías de evaluación: pertinencia, claridad, relevancia y suficiencia. La calificación se realizará mediante una escala de cuatro puntos (1 a 4), acompañada de descriptores cualitativos que explican el significado de cada nivel de valoración. Esta información permitirá estimar el grado de acuerdo entre jueces y determinar los ajustes necesarios en la redacción o formulación de los ítems antes de la aplicación piloto del instrumento.

Una vez analizados los ítems por el panel de jueces expertos, se registrarán sistemáticamente todas las sugerencias recibidas. Los ajustes correspondientes se realizarán en función de criterios tanto cuantitativos como cualitativos. Posteriormente, y de ser necesario, se llevarán a cabo reuniones de seguimiento con los jueces para obtener una reevaluación de los ítems modificados. Este proceso tiene como finalidad consolidar el instrumento de medición e implementar las mejoras identificadas, garantizando así su validez de contenido. Para evaluar el grado de acuerdo entre jueces, se utilizará la fórmula de la V de Aiken (Rodríguez et al., 2016). Esta fórmula interpreta V como un valor que cuantifica el acuerdo entre jueces considerando como un indicador importante el promedio de calificaciones (\bar{X}), pero toma en cuenta además la calificación mínima posible (l) y el rango de calificaciones posibles ($k = \text{máx.} - \text{mín.}$) mediante la fórmula:

$$v = \frac{\bar{x} - l}{k}$$

Complementariamente, se calcularán intervalos de confianza (IC) asimétricos para la V de Aiken. Los intervalos de confianza constituyen una herramienta estadística fundamental que permite estimar un rango de valores dentro del cual se encuentra el verdadero parámetro poblacional con un determinado nivel de confianza (generalmente 95%). En el contexto de la validación por jueces expertos, los IC para la V de Aiken proporcionan información crucial sobre la precisión y estabilidad de la estimación del coeficiente de Omega de McDonald. Mientras que la V de Aiken puntual nos da una estimación del grado de acuerdo, los IC nos indican la variabilidad esperada.

La relación entre la V de Aiken y sus intervalos de confianza es de complementariedad metodológica: la V proporciona la estimación puntual del acuerdo, mientras que los IC establecen los límites probables dentro de los cuales se ubicaría esta estimación en la población de posibles jueces. Para el cálculo de los intervalos de confianza asimétricos se utilizarán las siguientes fórmulas (Rodríguez et al., 2016):

$$L = \frac{2nkV + z^2 - z\sqrt{4nkV(1 - V) + z^2}}{2(nk + z^2)}$$

$$U = \frac{2nkV + z^2 + z\sqrt{4nkV(1 - V) + z^2}}{2(nk + z^2)}$$

Donde L y U son los límites inferior y superior del IC para la V respectivamente, Z es el valor en distribución estándar (1.96 para un 95% de confianza) y n es el número de jueces. Si bien no existe un consenso absoluto respecto al valor del límite inferior del IC de la V que puede considerarse como aceptable, la evidencia metodológica indica que 0.70 constituye un criterio adecuado (Rodríguez et al., 2016), ya que no permite que sea valorado como representativo un

ítem cuando el acuerdo entre jueces no es lo suficientemente alto. Este valor asegura que incluso en el escenario más conservador (límite inferior del IC), el acuerdo se mantiene en niveles psicométricamente aceptables.

Este enfoque mixto, que combina el análisis cualitativo de las sugerencias de los expertos con el análisis cuantitativo del acuerdo mediante la V de Aiken y sus intervalos de confianza, fortalece significativamente la evidencia de validez de contenido del instrumento y proporciona una base sólida para la toma de decisiones sobre la retención, modificación o eliminación de los ítems que conformarán la escala definitiva.

3.4.4 Entrevistas cognitivas.

Las entrevistas cognitivas se llevarán a cabo de forma presencial con el objetivo fundamental de evaluar la validez referida al proceso de respuesta, complementando así la validez de contenido obtenida mediante el juicio de expertos. Este método permite obtener evidencia directa sobre los procesos mentales que los participantes utilizan para comprender, recuperar información y formular una respuesta a cada ítem, siendo esencial para identificar problemas de comprensión, ambigüedades semánticas o dificultades en la toma de decisiones que podrían comprometer la validez de las puntuaciones (Benítez y Padilla, 2014; Willis, 2015). La aplicación será realizada por los investigadores a una submuestra de diez estudiantes de entre 7 y 10 años (cinco mujeres y cinco hombres) para garantizar equidad de género. Cada entrevista contemplará la aplicación de tres preguntas del cuestionario, con el propósito de evitar la fatiga cognitiva en los participantes (Jeong, 2022). La entrevista será realizada por dos investigadoras: una asume el rol principal, registrando manualmente las respuestas, interactuando de forma constante con el entrevistado y siguiendo un protocolo estandarizado (Anexo 1); la segunda cumple una función de

apoyo, toma apuntes digitales, controla el tiempo y, previa autorización, realizará una grabación de audio para facilitar el análisis posterior. Esta grabación servirá como respaldo para una triangulación más fina de los datos cualitativos. Posteriormente, los investigadores contrastan los resultados, comentarios y puntuaciones hasta alcanzar un consenso.

Cada ítem será evaluado durante las entrevistas en cuatro dominios clave: comprensión semántica y lingüística, comprensión de la tarea y estructura, comprensión conceptual (validez de constructo) y *engagement* y pertinencia percibida (Aracena et al., 2024). La evaluación de estos dominios se realizará mediante un sistema dual que combina una escala cuantitativa con una guía cualitativa detallada (Tablas 5 y 6).

La evaluación de estos cuatro dominios no es arbitraria, sino que se articula para mitigar fuentes específicas de error en la medición. La comprensión semántica y lingüística evalúa la validez de la representación verbal del ítem, asegurando que el vocabulario y la sintaxis sean accesibles. La comprensión de la tarea y estructura verifica que el formato de respuesta (selección múltiple) no constituya una barrera metodológica. La comprensión conceptual se enfoca en la validez del constructo, determinando si el estudiante asocia la situación presentada con el fenómeno de curiosidad epistémica que se pretende medir. Finalmente, el *engagement* y pertinencia percibida explora la validez motivacional, crucial para asegurar que la tarea elicite una respuesta genuina y no una elección al azar producto del desinterés o la falta de familiaridad con el contexto matemático presentado.

1. Escala de Valoración Cuantitativa

Para estandarizar la evaluación de cada dominio, se utilizará una escala Likert de 5 puntos que corresponde a lo ilustrado en la Tabla 5 que asigna una puntuación global al desempeño del ítem. En donde se especifica los tipos de cambios que se deben realizar, según la puntuación que obtenga cada ítem.

Tabla 5

Escala de valoración cuantitativa para entrevistas cognitivas

ETIQUETA	PUNTUACIÓN	DEFINICIÓN
Totalmente de acuerdo	5	El ítem cumple excelentemente. La evidencia es clara y fuerte en todos los aspectos (por ejemplo, los 4 indicadores son abordados de manera sobresaliente). El ítem no requiere modificaciones
De acuerdo	4	El ítem cumple bien con los indicadores. La evidencia es mayoritariamente positiva. El ítem requiere menores ajustes.
Neutral	3	El ítem cumple de manera aceptable pero no óptima con los indicadores. El ítem requiere ajustes importantes.
En desacuerdo	2	El ítem cumple deficientemente con los indicadores. El ítem requiere un ajuste estructural.
Totalmente en desacuerdo	1	El ítem no cumple en absoluto con los indicadores. El ítem debe ser rediseñado completamente.

Nota. Elaborada por los autores para esta investigación

2. Guía de Observación y Análisis Cualitativa

Paralelamente a la escala numérica, las investigadoras utilizarán una matriz de evaluación cualitativa (Tabla 6) para cada ítem. Esta guía operacionaliza los cuatro dominios a través de preguntas orientadoras para formular al participante y una lista de indicadores observables específicos. Esta herramienta tiene como propósito fundamentar de manera objetiva la puntuación asignada y recabar comentarios detallados para la posterior revisión de los ítems.

Tabla 6

Matriz de Evaluación Cualitativa para Entrevistas Cognitivas

DOMINIO A EVALUAR	INDICADORES OBSERVABLES	PUNTAJE 1-5	COMENTARIOS
Comprensión semántica y lingüística. ¿Hay alguna palabra que no sepas qué significa? ¿Explicame qué dice acá? (Apuntar a el enunciado no la consigna)	(1) El niño no pregunta el significado de ninguna palabra. (2) El niño parafrasea la pregunta correctamente usando sinónimos adecuados. (3) Su expresión facial no denota confusión al leer el ítem. (4) No necesita que se le repita la pregunta de forma adicional.		
Comprensión de la tarea y estructura	(1) Comprende que debe elegir una sola opción de respuesta. (2) Sabe cómo marcar su respuesta (rodear, tachar, etc.) sin ayuda. (3) Entiende la instrucción general de la prueba (ej: "lee y elige la opción que más se parezca a ti"). (4) El formato de selección múltiple (A-B-C-D) le es familiar y manejable.		
Criterio de observación			
Comprensión conceptual (validez de constructo) 1. ¿Qué tenías que hacer en esta tarea? 2. ¿Por qué elegiste esa respuesta? 3. ¿ Por qué no marcaste otra alternativa?	(1) Su explicación verbal del ítem se alinea con el constructo de curiosidad epistémica que se pretende medir (ej: "esto es sobre querer aprender cosas nuevas"). (2) La opción que eligió coincide con la justificación que da (validez interna de la respuesta). (3) No se distrae con elementos accesorios o irrelevantes del ítem.		
Engagement y pertinencia percibida 2. FAMILIAR ESTÁ ASOCIADA AL DESCRIPTOR DE CURIOSIDAD. NO DE LOS PIRATAS 3. observar. 4. registrar tiempo.	(1) Mantiene la atención durante la lectura y respuesta del ítem. (2) Verbaliza que la situación presentada le es familiar o posible (ej: "eso me ha pasado en el cole"). (3) Muestra interés genuino por el contenido de la pregunta (ej: sonríe, se queda pensativo). (4) El tiempo que tarda en responder es apropiado (ni inmediato ni excesivamente largo por distracción).		

Nota. Elaborada por los autores para esta investigación

3.4.5 Procedimiento de análisis y revisión de ítems.

Una vez finalizadas las entrevistas cognitivas, se procederá a un análisis sistemático de los ítems centrado en la identificación de oportunidades de mejora. El proceso se basará en la integración de los datos cuantitativos y cualitativos recopilados, priorizando la refinación del instrumento sobre la eliminación prematura de ítems.

El análisis se llevará a cabo mediante la consolidación de todos los puntajes asignados por los entrevistados en una hoja de cálculo. Para cada ítem, se calculará un promedio general y un promedio por cada uno de los cuatro dominios de evaluación (Comprensión semántica y lingüística, Comprensión de la tarea y estructura, Comprensión conceptual, y Engagement y pertinencia percibida). El criterio de decisión principal será el siguiente:

Se procederá a la revisión y refinamiento de cualquier ítem que obtenga un promedio general o un promedio en uno o más dominios inferiores a 4.0. El análisis cualitativo de los comentarios y observaciones registradas en las matrices será fundamental para diagnosticar la causa específica de la debilidad y guiar la modificación correspondiente. Por ejemplo, un promedio bajo en el dominio "Comprensión semántica" junto con comentarios sobre vocabulario complejo, indicará la necesidad de simplificar el lenguaje del ítem. Este enfoque de mejora continua es una práctica estándar en el desarrollo de instrumentos, ya que permite capitalizar la inversión inicial en el diseño de los ítems (Boateng et al., 2018).

Solo se considerarán para su retención sin cambios aquellos ítems que alcancen un promedio general y por dominio igual o superior a 4.0, y cuyos comentarios cualitativos no evidencian ambigüedades o problemas de interpretación.

La aplicación de este criterio será realizada de manera independiente por los dos investigadores principales, quienes contrastarán sus análisis en una sesión de consenso con un tercer investigador. En esta sesión, se discutirán las evidencias numéricas y cualitativas de cada ítem para decidir de manera unánime la naturaleza específica de las modificaciones requeridas. Toda esta información será documentada en una matriz de decisión final, la cual registrará para cada ítem los promedios, un resumen de los temas cualitativos clave y la justificación para cada modificación propuesta, asegurando la total transparencia y trazabilidad del proceso de revisión.

3.4.6 Aplicación Piloto

La versión refinada del instrumento, resultante del proceso de validación de contenido y entrevistas cognitivas, será aplicada en formato piloto a una muestra de entre 50 a 90 estudiantes. Esta fase tiene como objetivo principal evaluar el comportamiento psicométrico de los ítems en

condiciones de campo y verificar los supuestos estadísticos necesarios para el análisis con α de Cronbach. El tamaño de la muestra piloto se determina en base a estándares metodológicos que recomiendan un rango de 5 a 10 participantes por ítem, o un mínimo absoluto de 100 a 200 casos para garantizar la estabilidad de las estimaciones (Boateng et al., 2018; Fidell y Tabachnick, 2019), por lo que una muestra cercana a los 100 participantes se considera rigurosa para el instrumento de 8 ítems.

El análisis de los datos piloto se realizará en dos etapas secuenciales. Primero, se ejecutará un análisis descriptivo e ítem para examinar la calidad preliminar de cada reactivo. Esto incluirá el cálculo de la media, la desviación estándar y la correlación ítem-total corregida. Una desviación estándar muy baja (< 0.5) sugeriría poca variabilidad en las respuestas, mientras que un índice de discriminación (correlación ítem-total) inferior a 0.30 indicaría que el ítem no está contribuyendo adecuadamente a medir el constructo común (Field, 2018). Los ítems que no cumplan con estos criterios serán marcados para una revisión cualitativa que evalúe su posible reformulación o eliminación. Y en segundo lugar se realizará el análisis psicométrico clásico que consta del uso del Alfa de Cronbach. Este proceso evalúa la consistencia interna de la escala, verificando el grado de homogeneidad y coherencia con que sus ítems miden el constructo unidimensional de curiosidad epistémica. El valor esperado para el coeficiente es de 0.70 o superior, considerado el umbral mínimo aceptable que avalaría la confiabilidad del instrumento para su uso en investigación (George y Mallery, 2019). Un resultado dentro de este rango indicaría que las mediciones son estables y que los ítems contribuyen de manera fiable a la medición global del constructo.

3.4.7 Análisis psicométrico clásico

Fiabilidad

La fiabilidad es una característica esencial de todo instrumento de medición y se refiere a la capacidad de este para ofrecer resultados consistentes y estables a lo largo del tiempo y en distintas condiciones. En el contexto del instrumento que estamos desarrollando para medir la curiosidad epistémica en matemática en estudiantes de cuarto básico, la fiabilidad indica qué tan confiables y precisas son las puntuaciones obtenidas, es decir, si el instrumento mide de forma coherente el constructo en estudio sin que los resultados se vean afectados por errores aleatorios.

Se priorizará la fiabilidad para garantizar su reutilización. Un instrumento que haya superado la fase de validez y fiabilidad de forma precisa y correcta podría ayudar al investigador a medir las variables estudiadas y a obtener resultados precisos de dicho análisis. Además, este hallazgo respalda otras investigaciones, lo que permite realizar el análisis de los instrumentos de forma precisa y correcta (Ahmad y Siew, 2021).

En este estudio, el análisis descriptivo del desempeño, índice de dificultad de cada ítem y confiabilidad interna del instrumento se determinará exclusivamente a través del cálculo del coeficiente Alfa de Cronbach. Este coeficiente es el estimador más ampliamente utilizado y aceptado para evaluar la consistencia interna de escalas tipo Likert, ya que cuantifica el grado en que los ítems que componen una escala se correlacionan entre sí, midiendo así un constructo unidimensional (Cronbach, 1951; Tavakol y Dennick, 2011).

Siguiendo los criterios psicométricos convencionales establecidos en la literatura, se interpretarán los resultados de la siguiente manera (George y Mallery, 2003): Valores iguales o superiores a

0.70 se considerarán como evidencia de una consistencia interna aceptable para fines de investigación en ciencias sociales. Los valores comprendidos entre 0.80 y 0.90 se considerarán buenos, lo que sugiere una alta homogeneidad de los ítems. Valores por encima de 0.90 se considerarán excelentes. El alfa de Cronbach nos ayudará con los siguientes indicadores.

- **Análisis Descriptivo:** Evaluación de la distribución y variabilidad de las respuestas para cada ítem y la escala total, identificando problemas como baja dispersión (desviación estándar < 0.5) que limiten la capacidad discriminativa.
- **Índice de Dificultad/Nivel de Aceptación:** Cálculo del promedio de respuesta por ítem para detectar reactivos extremos (con muy alto o muy bajo acuerdo), los cuales aportan poca información útil para diferenciar entre los participantes.
- **Confiabilidad Interna:** Estimación del coeficiente Alfa de Cronbach para la escala completa, considerando un valor ≥ 0.70 como aceptable (Bernstein y Nunnally, 1994). Se analizará además el impacto de eliminar cada ítem en la consistencia interna global

La obtención de un Alfa de Cronbach satisfactorio respalda la precisión de las mediciones realizadas con el instrumento y fortalece la validez de las interpretaciones y conclusiones que se derivarán de su aplicación en esta investigación.

3.4.8 Programa de datos que se utilizara en el estudio

Para el análisis estadístico del instrumento se utilizarán los programas Statistical Package for the Social Sciences (SPSS), herramienta ampliamente reconocida y empleada en la investigación psicométrica. SPSS es un software de análisis estadístico de fácil manejo que permite

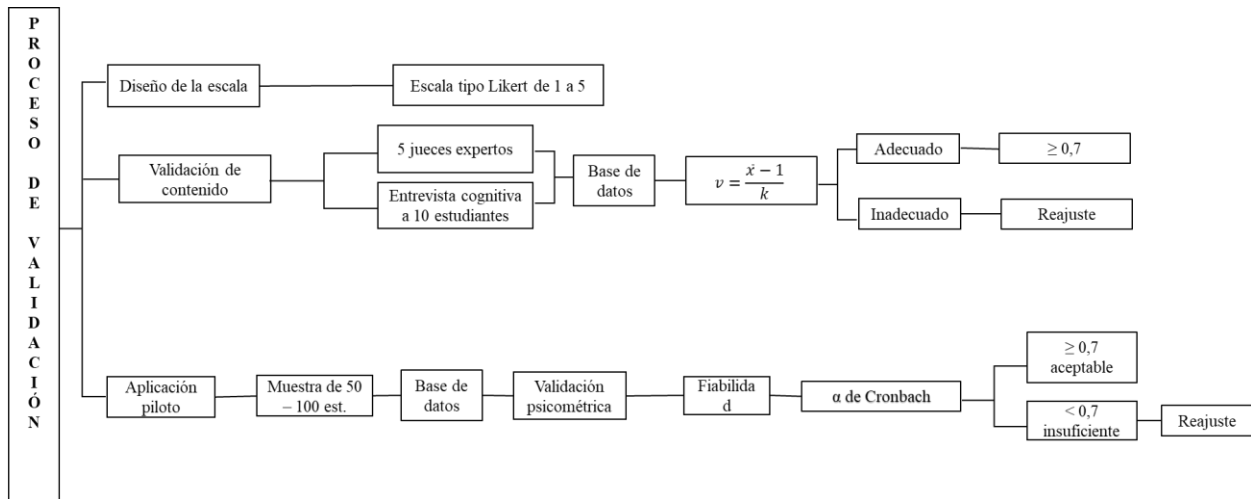
realizar cálculos descriptivos, análisis de confiabilidad y correlaciones, facilitando la exploración inicial de los datos y la interpretación de resultados.

El uso de SPSS permite aprovechar las fortalezas de sus programas facilitando la gestión y análisis preliminar de los datos en un entorno amigable. Este nos garantiza un análisis riguroso, preciso y adecuado al contexto educativo y cultural del sistema escolar chileno, siguiendo las recomendaciones metodológicas para estudios psicométricos en Latinoamérica (Aracena et al., 2024). Así, logra contribuir complementariamente al desarrollo y validación del instrumento, asegurando la calidad y robustez de los resultados obtenidos.

A continuación, se presenta la Figura 3 un resumen del proceso de validación

Figura 3

Esquema resumen del proceso de validación paso a paso.



3.5 CONSIDERACIONES ÉTICAS

La investigación respetará los principios éticos establecidos en la normativa chilena e internacional para estudios con menores de edad (Ministerio de Salud de Chile, 2006). Uno de los

principios básicos de la ética de la investigación afirma que todo ser humano debe ser respetado en su decisión de participar o negarse a participar de una investigación, y ello mediante un proceso de consentimiento previo, libre e informado (Bahamondes et al., 2017). Se solicitará el consentimiento informado a padres y apoderados y se realizará el proceso de asentimiento, el cual debe ser coherente con el derecho reconocido en la Convención sobre los Derechos del Niño en su artículo 12, que dispone: “Los Estados garantizarán al niño que esté en condiciones de formarse un juicio propio, el derecho de expresar su opinión libremente en todos los asuntos que le afecten, teniéndose debidamente en cuenta sus opiniones, en función de su edad y madurez” (Bahamondes et al., 2017), garantizando la confidencialidad y el anonimato de las respuestas, e informando a los participantes que pueden retirarse del estudio en cualquier momento sin repercusiones, dichos documentos deben ser aprobados por un comité de ética de la universidad de Concepción para ser entregados a padres y/o apoderados y estudiantes.

La aplicación del instrumento se realizará únicamente en establecimientos que otorguen autorización formal mediante carta de compromiso institucional. El estudio será conducido con estricto apego a criterios de respeto, dignidad y resguardo de la integridad emocional y cognitiva de los niños y niñas participantes. (MINSAL, 2006)

IV. ANÁLISIS Y RESULTADOS

A continuación, se presentan los resultados obtenidos correspondientes a la primera etapa del proceso de validación de contenido del instrumento diseñado, derivada de la aplicación de entrevistas cognitivas.

4.1 RESULTADOS DERIVADOS DE LAS ENTREVISTAS COGNITIVAS

Este apartado detalla los resultados de las entrevistas cognitivas aplicadas a una muestra de diez estudiantes, 5 hombres y 5 mujeres, con edades comprendidas entre los 7 y 10 años, de la provincia del Biobío, donde los ítems 1, 2, 3, 4, 5, 6 y 7 se realizaron 3 entrevistas por ítem a excepción del ítem 8 donde se realizaron 6 entrevistas por ítem. Los participantes no presentaban necesidades educativas especiales (NEE). El análisis se fundamentó en cuatro criterios de evaluación: comprensión semántica y lingüística, comprensión de la tarea, comprensión conceptual, y engagement y pertinencia percibida. Cada criterio fue evaluado por los investigadores utilizando una escala Likert del 1 al 5, donde 1 indica que el ítem debe ser rediseñado completamente y 5 que cumple de forma correcta y no requiere modificaciones (tabla 5 y tabla 6).

Para la interpretación de los datos, se estableció: una puntuación promedio de 5 denota un ítem óptimo; una puntuación de 4 señala la necesidad de ajustes menores para optimizar la claridad o precisión; y una puntuación de 3 o inferior evidencia problemas sustanciales que requieren ajustes mayores o una reformulación profunda.

La Tabla 7 presenta los datos recopilados los cuales fueron consolidados en una hoja de cálculo para obtener los puntajes promedio por criterio para cada ítem.

Tabla 7

Resultados obtenidos de las entrevistas cognitivas según criterios de evaluación

Ítem	Cantidad de entrevistados	Comprensión semántica y lingüística	Comprensión de la tarea	Comprensión conceptual	Engagement y pertinencia percibida
1	3	4,0	5,0	5,0	4,0
2	3	4,3	5,0	4,3	4,7
3	3	4,3	5,0	3,7	4,3
4	3	4,0	5,0	5,0	4,0
5	4	4,3	5,0	4,3	4,5
6	3	4,0	5,0	5,0	4,7
7	4	5,0	5,0	4,0	4,5
8	7	3,3	4,6	3,6	4,4

A partir de los resultados consignados en la Tabla 7, se observa un desempeño general favorable del instrumento. Del total de ocho ítems evaluados, seis (que representan el 75%) presentaron puntuaciones entre cuatro y cinco en la mayoría de los criterios, lo que indica un

cumplimiento satisfactorio o excelente de los indicadores evaluados y sugiere una buena calidad inicial del cuestionario.

Sin embargo, el análisis permitió identificar dos ítems con puntuaciones inferiores a cuatro en criterios específicos, lo que evidencia un desempeño meramente aceptable y no óptimo. En particular, el Ítem 8 es el que presenta mayores desafíos, con puntuaciones por debajo de cuatro en tres de los cuatro criterios, especialmente en Comprensión semántica y lingüística (3,3) y Comprensión conceptual (3,6). Asimismo, el Ítem 3 muestra una debilidad específica en Comprensión conceptual (3,7).

En función de este diagnóstico, se procedió a un examen exhaustivo de los ítems problemáticos con el fin de implementar las adecuaciones pertinentes para cada criterio, las cuales se especifican en el siguiente apartado.

4.1.1 Comprensión semántica y lingüística

Este criterio valora la competencia lingüística de los estudiantes a través de tres dimensiones interrelacionadas: la claridad, la precisión y la adecuación del lenguaje empleado en los ítems. Su aplicación permite determinar si los participantes comprenden de manera clara el vocabulario y las estructuras lingüísticas utilizadas. Asimismo, se verifica su habilidad para reformular los enunciados, conservando de manera fiel el sentido original, y se constata la ausencia de indicios de ambigüedad o confusión que pudieran obstaculizar una interpretación correcta ya sea en la pregunta o alternativas de cada ítem.

Los resultados obtenidos del criterio comprensión semántica y lingüística evidencia que un 87,5% obtuvo entre cuatro y cinco en la puntuación los cuales según la escala likert (Tabla 5) son ítems que cumplen de forma sobresaliente con los indicadores propuestos como lo es en el caso de

aquellos ítems que obtuvieron 4 y aquellos que obtuvieron entre 4,7 y 4,0 requieren modificaciones menores en aspectos específicos para llegar a un nivel óptimo, sin embargo existe un 12,5% que obtuvo una puntuación inferior a 4,0 por lo que requiere mayor atención en este resultado corresponde al ítem 8 (Figura 4). Las entrevistas permitieron detectar tres problemas críticos en el ítem original. Primero, se evidenció que la Tabla presentaba una ilustración poco clara para los estudiantes, lo que dificultó su lectura y comprensión para responder el ítem. Segundo, una desconceptualización del término "conclusión", donde los estudiantes mostraron incapacidad para definir o explicar este sustantivo abstracto, obstaculizando la comprensión misma de la pregunta. Tercero, se identificó una interpretación errónea del adjetivo "inquieto", que los participantes asociaron consistentemente con una connotación conductual de "desorden" en lugar de la dimensión emocional de intranquilidad que se pretendía evaluar.

Estos hallazgos exigieron una reformulación del ítem orientada a optimizar las tres dimensiones del criterio. En la dimensión de claridad, se reemplazaron las etiquetas genéricas "Resultado 1, 2, 3" por la estructura ordinal explícita "1º movimiento, 2º movimiento y 3º movimiento", eliminando la ambigüedad temporal. En la dimensión de adecuación del lenguaje empleado, se incorporaron dos ajustes fundamentales: El sustantivo "conclusión" fue reemplazado por "decisión" y el adjetivo "inquieto" por "ansiedad" ajustándose con mayor exactitud a la lógica de la acción narrativa para una mejor comprensión de los y las estudiantes.

Figura 4

Versión original del ítem 8 utilizado en el "instrumento para la evaluación de entrevista cognitiva"

Después de agitar cada botella 3 veces, Tomás obtuvo estos resultados:

Resultados	Botella 1 (perlas de oro)	Botella 2 (perlas de oro)	Botella 3 (perlas de oro)
Resultado 1	9	3	3
Resultado 2	4	6	3
Resultado 3	3	5	4

Tomás eligió la botella 1, porque una vez contó 9 perlas de oro.

¿Cómo te sientes con la conclusión de Tomás?

- a) Me siento bien. La botella 1 mostró el número mayor de perlas de oro.
- b) Emocionado, me gustaría ver si vuelven a salir 9 perlas de oro.
- c) Inquieto. Necesito ver si vuelve a sacar 9.
- d) Me da igual. Si nos equivocamos es culpa de él.

En síntesis, la triangulación de datos permitió diagnosticar y rectificar las carencias lingüísticas del ítem 8. Las modificaciones implementadas se alinean con el principio de accesibilidad lingüística, asegurando que la evaluación mida correctamente la dimensión de comprensión semántica y lingüística. La versión final conserva el objetivo pedagógico original mientras elimina los ruidos semánticos que impedían su correcta interpretación.

4.1.2 Comprensión de la tarea y estructura

El criterio de comprensión de la tarea y estructura evaluó la competencia de los estudiantes para descifrar la demanda general de la actividad, lo que implicaba reconocer la necesidad de seleccionar una única opción de respuesta y señalarla de manera autónoma, sin requerir asistencia externa. Asimismo, este criterio valoró el grado de familiarización del alumnado con la estructura protocolar de los ítems de opción múltiple (alternativas a, b, c, d). Los resultados cuantitativos reflejan un desempeño excepcional en este ámbito, constituyéndose como el criterio mejor

evaluado de todo el instrumento. La práctica totalidad de los ítems, un 87.5%, obtuvo la puntuación máxima de 5.0, lo que indica que los estudiantes comprendieron sin dificultad las instrucciones y supieron interactuar con el formato de respuesta de manera correcta e independiente. La única excepción fue el Ítem 8, que a pesar de ser el de menor puntuación en este criterio, consiguió un 4.6, una calificación que se sitúa aún en el umbral de cumplimiento satisfactorio y que solo sugiere la necesidad de ajustes menores. Por consiguiente, en nuestro rol de investigadores, se decidió implementar un ajuste estructural en el enunciado: “Tomás eligió la botella 1, porque una vez contó 9 perlas de oro”. Las modificaciones consistieron en aumentar el tamaño y la intensidad (oscurecimiento) de la tipografía, así como en aislar visualmente el enunciado del cuadro y de la pregunta. Esta intervención se justificó tras observar que los estudiantes tendían a omitir la información del enunciado, lo que obstaculizaba la comprensión y ejecución de la tarea propuesta. Este resultado globalmente positivo demuestra que la estructura general del instrumento es adecuada para la población objetivo y que el formato de selección múltiple es comprensible y aplicable para niños del rango de edad de 7 a 10 años, validando así el diseño de la tarea en su conjunto.

4.1.3 Comprensión conceptual

Este criterio evalúa la capacidad de los estudiantes para comprender los constructos subyacentes de curiosidad epistémica de tipo I y tipo D, según la tipología de Litman, que el instrumento busca medir. La evaluación consideró si los participantes podían explicar el sentido profundo del ítem, justificar su elección mediante un razonamiento lógico basado en las evidencias presentadas, diferenciar la información relevante de la irrelevante y demostrar su capacidad para rechazar críticamente las alternativas incorrectas.

Los resultados cuantitativos para este criterio revelan un panorama mixto. Si bien el 62.5% de los ítems (5 de 8) obtuvo puntuaciones entre 4.0 y 5.0, indicando una comprensión conceptual generalmente adecuada, un 37.5% (3 de 8) presentó puntuaciones inferiores a 4.0. Los ítems 3 (3.7) y 8 (3.6) emergen como los más críticos, evidenciando dificultades sustanciales en la asimilación de los conceptos objetivos. El análisis cualitativo sugiere que estos déficits no se originaron primordialmente en una falta de capacidad cognitiva de los estudiantes, sino que fueron precipitados por las barreras lingüísticas identificadas en el criterio de comprensión semántica. La incomprensión de terminología clave imposibilita que los alumnos accedieran plenamente a la demanda conceptual, comprometiendo su capacidad para elaborar justificaciones sólidas y refutar alternativas de manera crítica.

El análisis de los resultados derivados de las entrevistas cognitivas identificó deficiencias críticas en la comprensión conceptual de los ítems 3 y 8, lo que motivó una reformulación profunda de ambos.

Para el ítem 3 (Figura 5), el problema central residía en su incapacidad para discriminar entre una elección basada en la preferencia y una fundamentada en un proceso de indagación crítica. Una barrera significativa fue la mención de "ver la tabla de Juanito", ya que el ítem en sí no incluía ninguna tabla. Esta referencia a un elemento inexistente generó confusión y dificultó que los estudiantes entendieran la situación presentada. La alternativa más inadecuada, "Ver la tabla de Juanito y anotar todo", era conceptualmente débil. Si bien esta opción acercaba al estudiante a la evidencia, se limitaba a una acción pasiva de registro que no exigía comprender la necesidad de evaluar la validez o el origen de la afirmación presentada. Esto permitía que los estudiantes seleccionaran una respuesta correcta sin necesariamente activar un razonamiento

basado en la curiosidad epistémica de tipo I, la cual se caracteriza por una sensación de interés que impulsa a buscar información faltante.

Figura 5

Versión original del ítem 3 utilizado en el “instrumento para la evaluación de entrevista cognitiva”

El pirata Juanito agitó las botellas y se emocionó al decir: La botella 3 es la mejor.

¿Qué te gustaría hacer?

- a) Creerle y dar por terminado el juego.
- b) Ver la tabla de Juanito y anotar todo.
- c) Agitarlas yo también, parece divertido ver qué pasa.
- d) Nada, pensar esto es aburrido.

La versión mejorada del ítem supera esta limitación mediante una modificación estratégica en la redacción de la alternativa clave, que evoluciona a "Preguntar al pirata Juan y anotar como lo hizo". Este cambio, aparentemente sutil, es conceptualmente trascendental. Al incorporar el verbo "preguntar" y el interrogante "cómo lo hizo", el ítem ya no evalúa sólo el deseo de consultar datos, sino la comprensión activa de que el conocimiento se construye sobre métodos y procedimientos verificables. Esta reformulación evalúa directamente la capacidad del estudiante para reconocer que una afirmación, especialmente proveniente de una figura de autoridad, requiere una justificación sobre su proceso de generación. De este modo, el ítem se transforma de una mera consulta sobre preferencias en una herramienta capaz de medir el escepticismo informado y la comprensión conceptual de la relación entre evidencia, método y conclusión.

En el mismo sentido, el ítem 8 (Figura 4) fue objeto de una reformulación debido a que su estructura inicial obstaculiza la aprehensión conceptual y comprometía la validez de la evaluación.

La principal modificación consistió en reemplazar la notación simbólica de la tabla por una representación pictórica, con el fin de hacer más intuitiva la naturaleza de los datos. Adicionalmente, se introdujeron divisorias verticales para establecer una clara distinción entre los resultados de los diferentes movimientos, optimizando la legibilidad y el proceso de análisis por parte de los estudiantes.

Estos ajustes simplificaron la estructura de la tabla de datos para facilitar su interpretación, teniendo en cuenta la edad de la población objetivo. En conjunto, las modificaciones convierten al ítem en una herramienta válida para evaluar la comprensión de conceptos estadísticos básicos, como la representatividad de los datos.

4.1.4 Engagement y pertinencia percibida

El criterio de engagement y pertinencia percibida evaluó la capacidad del instrumento para captar y mantener el interés de los estudiantes mediante situaciones contextualizadas que resultan comprensibles, cercanas a su experiencia cotidiana y coherentes con su marco de referencia. Este criterio consideró específicamente el nivel de atención demostrado durante la lectura de los ítems, la ausencia de distracciones y la capacidad de responder dentro de un tiempo adecuado, reflejando así una conexión efectiva entre el contenido presentado y la disposición cognitiva y motivacional de los participantes.

Los resultados obtenidos en este criterio (Tabla 7) son notablemente positivos, constituyéndose como los más altos y consistentes de toda la evaluación. El 100% de los ítems logró puntuaciones iguales o superiores a 4.0, con una clara tendencia hacia la excelencia: el 62.5% de los ítems (5 de 8) alcanzó la puntuación máxima de 5.0, mientras que el 37.5% restante (3 de 8) se ubicó en el rango de 4.0 a 4.7. Este desempeño uniformemente alto indica que las narrativas

y escenarios planteados especialmente aquellos que incorporan elementos lúdicos como piratas, tesoros y descubrimientos consiguieron resonar efectivamente con la población objetivo, generando una experiencia de evaluación percibida como significativa y relevante.

Cabe destacar que, incluso en los ítems que presentaron dificultades en otros criterios como los ítems 3 y 8, que mostraron problemas de comprensión conceptual y semántica las puntuaciones en engagement se mantuvieron en niveles satisfactorios (4.3 y 4.4, respectivamente). Este fenómeno sugiere que, si bien ciertos aspectos lingüísticos o conceptuales pueden haber obstaculizado la comprensión plena, el diseño contextual y narrativo de los ítems logró mantener la motivación y atención de los estudiantes. Este hallazgo es particularmente relevante, pues valida el diseño general del instrumento en términos de su capacidad para generar una experiencia de evaluación positiva y atractiva, sentando las bases para una aplicación efectiva en contextos educativos reales con población infantil. La consistencia en estos resultados refuerza la viabilidad del instrumento como herramienta diagnóstica, al demostrar que logra el equilibrio necesario entre rigor metodológico y adecuación psicoeducativa.

4.2 RESULTADOS DE VALIDACIÓN DEL CONTENIDO MEDIANTE EL JUICIO DE EXPERTOS

El proceso de validación de contenido mediante el juicio de expertos constituye una fase metodológica fundamental en el desarrollo de instrumentos de investigación, ya que garantiza que los ítems que lo componen sean conceptualmente sólidos, claros y pertinentes para el constructo que se pretende medir. Este procedimiento, llevado a cabo en el presente estudio, se estructuró en un diseño de dos rondas sucesivas, permitiendo no solo una evaluación inicial sino también una depuración y mejora sistemática del instrumento.

La primera ronda de validación proporcionó la valoración basal de los ocho ítems a través del coeficiente V de Aiken, analizando cuatro criterios esenciales: Claridad, Pertinencia, Relevancia y Suficiencia. Los resultados iniciales, recogidos en la Tabla 8, revelaron un nivel de aceptación generalmente bueno, aunque se identificaron oportunidades de mejora específicas. Por ejemplo, ítems como el 4 y el 1 mostraron valores tales como 0.770 y 0.792 respectivamente en algunos criterios, indicando la necesidad de refinamientos para optimizar su comprensión y alcance detalles que se abordan con mayor profundidad en los siguientes encabezados. Estos datos iniciales sirvieron como diagnóstico preciso para la fase de revisión.

Posteriormente, se llevó a cabo una fase de revisión y ajuste de los ítems, en la cual las observaciones y puntuaciones de los expertos fueron analizadas cualitativamente. Este análisis permitió reestructurar redacciones, precisar conceptos y ajustar el enfoque de aquellos ítems que no alcanzaron los niveles de excelencia requeridos, con el objetivo de fortalecer la validez de contenido global del instrumento.

Finalmente, se implementó una segunda ronda de evaluación esta vez solo con aquellos jueces que asignaron una puntuación insuficiente (entre 1 y 3), utilizando la versión modificada del instrumento. Los resultados finales, consignados en la Tabla 9, muestran el impacto positivo de los ajustes realizados. Se observa una mejora generalizada y consistente en los valores del coeficiente V de Aiken para la mayoría de los ítems y criterios, destacándose el significativo incremento en la claridad del ítem 3 (de 0.556 a 0.778) y una homogeneización hacia valores superiores en el criterio de Pertinencia. La convergencia de los valores hacia el máximo teórico ($V=1.0$) en esta ronda final corrobora la efectividad del proceso iterativo y confirma que el

instrumento alcanza unos estándares de calidad y validez de contenido óptimos para su aplicación en el campo de estudio.

4.2.1 Evaluación inicial de la validez de contenido por jueces expertos

En este apartado, se presenta el análisis de los resultados obtenidos en la primera ronda de evaluación por jueces expertos, utilizando el coeficiente V de Aiken. La Tabla 8 muestra los valores iniciales de validez para cada ítem en los cuatro criterios evaluados. Para un análisis integral, se ha calculado el promedio total del V de Aiken por ítem, el cual ofrece una visión global de su aceptación inicial.

Tabla 8

Valores iniciales y promedio por ítem con V de Aiken

Ítem	Claridad	Pertinencia	Relevancia	Suficiencia	Promedio (V)
1	0.833	0.667	0.889	0.778	0.792
2	0.889	0.944	1.000	0.917	0.938
3	0.556	0.889	1.000	0.933	0.845
4	0.800	0.778	0.722	0.778	0.770
5	0.722	0.889	0.833	0.889	0.833

6	0.778	0.833	0.889	0.667	0.792
7	0.667	0.833	0.889	0.778	0.792
8	0.778	0.889	0.944	1.000	0.903

Para la interpretación de los resultados obtenidos en la primera ronda de evaluación, se adoptaron los rangos de clasificación del coeficiente V de Aiken establecidos en la literatura especializada (Aiken, 1985; Penfield y Giacobbi, 2004). De acuerdo con Aiken 1985, un valor de $V \geq 0.90$ se considera de validez excelente, por lo que el ítem puede ser aceptado sin modificaciones. Un valor en el rango de $0.80 \leq V < 0.90$ indica una validez buena a muy buena, sugiriendo que el ítem es aceptable, pero podría optimizarse con pequeñas revisiones. Un valor entre $0.70 \leq V < 0.80$ se clasifica como validez aceptable, lo que significa que el ítem requiere revisiones moderadas para mejorar su calidad. Por último, un valor $V < 0.70$ se interpreta como validez baja, recomendándole que el ítem sea reformulado sustancialmente o eliminado.

A partir del diagnóstico obtenido en la primera ronda de validación, se procedió a una revisión sistemática de los ítems identificados como problemáticos. Este proceso no se limitó a ajustes superficiales, sino que implicó una reformulación fundamentada directamente en las observaciones cualitativas de los jueces expertos, con el objetivo de subsanar las deficiencias específicas en claridad, pertinencia, relevancia y suficiencia, modificaciones que son explicitadas en los siguientes 4 apartados.

4.2.2 Claridad

Este criterio de evaluación tuvo como objetivo valorar la precisión lingüística y la ausencia de ambigüedad en la redacción de cada ítem del instrumento. Un ítem se considera claro cuando es comprendido de manera precisa y unívoca por la población objetivo, condición fundamental para que las respuestas reflejen fielmente la dimensión de la curiosidad que se pretende medir, sin distorsiones derivadas de problemas de interpretación en el enunciado o en sus alternativas de respuesta. Los resultados obtenidos del criterio de claridad para el instrumento de medición fueron ordenados según la puntuación de cada juez y finalmente promediados con V de Aiken (Tabla 9)

Tabla 9

Resultados iniciales del grado de acuerdo según el criterio de evaluación de Claridad (V de Aiken)

Ítem	Juez 1	Juez 2	Juez 3	Juez 4	Juez 5	Juez 6	V de Aiken
1	4	4	3	4	3	3	0.83
2	4	4	4	4	3	3	0.89
3	3	3	3	3	2	2	0.56
4	4	3	3	4	NA	3	0.80
5	4	2	4	4	3	2	0.72

6	3	4	2	4	4	3	0.78
7	4	4	3	3	3	1	0.67
8	3	3	3	4	4	3	0.7

Los resultados iniciales de la validación agregar tabla por jueces expertos revelaron puntuaciones de claridad por debajo del nivel óptimo en tres ítems. En concreto, el ítem 3 obtuvo una valoración crítica de 0.56, clasificándose como de validez baja. Los revisores señalaron que su redacción era ambigua, especialmente en lo referente a las opciones de respuesta, y que emplea una terminología excesivamente técnica, lo que comprometía su comprensión uniforme. La observación más frecuente fue la necesidad de eliminar elementos contextuales externos que pudieran generar ambigüedad o confusión en el respondiente. Específicamente, en el caso del ítem 3, los jueces señalaron la ambigüedad de la mención a una "tabla" en la alternativa B, la cual no estaba presente ni definida en el enunciado principal de la pregunta. Esta referencia a un elemento ausente fue identificada como un factor crítico que comprometía la comprensión directa de la tarea requerida, forzando a los estudiantes a realizar inferencias no intencionadas para interpretar la opción de respuesta.

La corrección unánime propuesta y aplicada consistió en una reestructuración lingüística de dicha alternativa para asegurar que la pregunta fuera autónoma y auto explicativa. La alternativa original, que hacía referencia a una "tabla" de manera descontextualizada, fue reformulada para describir la acción o el criterio de manera clara y completa dentro del texto mismo de la opción. Por ejemplo, la alternativa genérica B del ítem 3 "Ver la tabla de Juanito y anotar todo" fue ser transformada en "Preguntar al pirata Juan y anotar como lo hizo", eliminando así la referencia

ambigua y encapsulando toda la información necesaria para que el respondiente comprenda la acción propuesta sin necesidad de elementos externos.

Asimismo, los ítems 5 y 7 presentaron valores de 0.72 y 0.67, respectivamente, lo que indicó la necesidad de refinar su estructura sintáctica y añadir aclaraciones para facilitar su comprensión. Por ejemplo, para el Ítem 5, una de las juezas (Jueza 6) sugirió modificar la alternativa “Agitarlas todas, me emociona mucho descubrir la más valiosa” por “Agitar todas las botellas, me emociona mucho descubrir la más valiosa”, con el fin de hacer el texto más explícito para el estudiantado. Respecto al Ítem 7, la misma jueza destacó la necesidad de garantizar que la distribución de las perlas fuera uniforme en el escenario descrito (Figura 6). Es decir, que todas las tablas contarán con la misma cantidad de perlas. En el caso de las botellas de Antonia, había 7 perlas en total, mientras que en las de Tomás existía una desigualdad (la botella 1 contenía 7 perlas y la botella 2, 8). Por consiguiente, se optó por estandarizar el escenario asignando nueve perlas a cada botella (Figura 7)

Figura 6

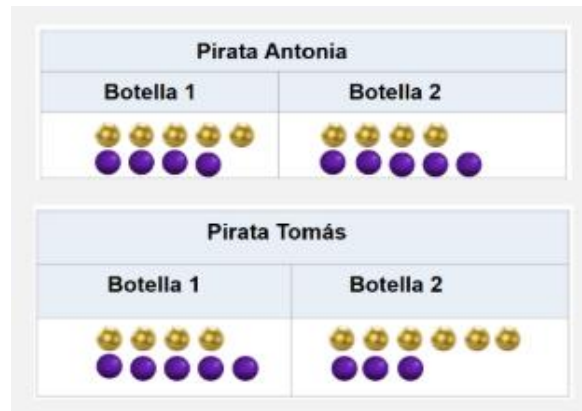
Versión original del ítem 7 utilizado en el “instrumento para la evaluación de entrevista cognitiva”

El diagrama muestra dos tablas de perlas. La primera tabla, para Antonia, tiene dos botellas: Botella 1 con 7 perlas (4 amarillas y 3 púrpuras) y Botella 2 con 7 perlas (3 amarillas y 4 púrpuras). La segunda tabla, para Tomás, tiene dos botellas: Botella 1 con 7 perlas (3 amarillas y 4 púrpuras) y Botella 2 con 8 perlas (5 amarillas y 3 púrpuras).

Pirata	Botella 1	Botella 2
Antonia	4 amarillas, 3 púrpuras	3 amarillas, 4 púrpuras
Tomás	3 amarillas, 4 púrpuras	5 amarillas, 3 púrpuras

Figura 7

Versión final del ítem 7 utilizado en el “instrumento para la evaluación de entrevista cognitiva”



Tras implementar estas mejoras, se convocó una reunión de seguimiento con los Jueces 1, 5 y 6 para presentarles los cambios realizados. Tras la exposición de las modificaciones por parte de los investigadores, los jueces aceptaron las mejoras y procedieron a reevaluar los ítems. Como resultado, las puntuaciones de claridad experimentaron un aumento significativo (ver Tabla 10): el ítem 3 pasó de 0.56 a 0.78; el ítem 5 incrementó de 0.72 a 0.78; y el ítem 7 mejoró de 0.67 a 0.72. Tras esta revisión, todos los ítems alcanzaron un nivel de validez satisfactorio, permitiendo así proceder con la fase de aplicación piloto del instrumento.

Tabla 10

Resultados finales del grado de acuerdo según el criterio de evaluación de Claridad (V de Aiken)

Ítem	Juez 1	Juez 2	Juez 3	Juez 4	Juez 5	Juez 6	V de Aiken
------	--------	--------	--------	--------	--------	--------	------------

1	4	4	3	4	4	3	0.89
2	4	4	4	4	3	3	0.89
3	3	3	3	4	4	3	0.78
4	4	3	3	4	4	3	0.83
5	4	2	4	4	3	3	0.78
6	3	4	2	4	4	3	0.78
7	4	4	3	3	3	2	0.72
8	3	3	3	4	4	3	0.78

4.2.3 Pertinencia

El criterio de pertinencia busca valorar el grado en que el contenido del ítem se alinea de manera directa y específica con la característica de la curiosidad epistémica (ya sea de interés o por privación) que pretende operacionalizar. Este criterio es fundamental, ya que asegura que cada ítem constituye una manifestación válida y representativa del constructo teórico que define a cada dimensión, garantizando así la validez de contenido del instrumento. Las puntuaciones de los jueces expertos fueron recopiladas y ordenadas según el ítem evaluado para finalmente obtener el promedio en la V de Aiken (Tabla 11)

Tabla 11

Resultados iniciales del grado de acuerdo según el criterio de evaluación de Pertinencia (V de Aiken)

Ítem	Juez 1	Juez 2	Juez 3	Juez 4	Juez 5	Juez 6	V de Aiken
1	4	4	3	2	3	2	0.67
2	4	4	3	4	4	4	0.94
3	4	4	3	4	4	3	0.89
4	4	4	3	4	1	4	0.78
5	4	4	3	4	4	3	0.89
6	4	3	2	4	4	4	0.83
7	4	4	3	4	4	2	0.89
8	4	4	3	4	4	3	0.89

La aplicación de los criterios establecidos a los promedios generales por ítem permitió una clasificación inicial de su validez. Los ítems 2 ($V=0.94$) y 8 ($V=0.89$) demostraron una validez excelente, ubicándose en el rango más alto. Esto indicó una aceptación unánime por parte del panel de jueces, por lo que no se consideraron necesarias modificaciones. Por su parte, los ítems 3 y 5 ($V=0.89$) se situaron en un nivel de validez de buena a muy buena. Si bien su promedio global era alto, un análisis detallado por criterios reveló la existencia de áreas específicas susceptibles de mejora.

Un caso particular fue el ítem 1, que inicialmente mostró una validez baja ($V=0.67$). Los expertos observaron que su contenido se solapaba significativamente con el del ítem 2, sin lograr

capturar un aspecto único y esencial del constructo, lo que puso en duda su pertinencia directa. Esta superposición se debía al parecido conceptual entre las subdimensiones “Motivación intrínseca para obtener información” y “Se activa al conocer algo nuevo”.

Para resolver esta discrepancia, se realizó una reunión de clarificación con el juez 4. En ella se explicó que, si bien podían parecer similares, el ítem 1 se centra en la necesidad interna de adquirir nuevo conocimiento (A1), mientras que el ítem 2 se enfoca en la curiosidad que surge como respuesta a una información novedosa explícita (A2). Esta diferencia fundamental se evidenciaba en el contexto de cada pregunta: el ítem 1 ("Tomás ya eligió su botella, ¿qué te gustaría hacer con las que quedan?") presenta un escenario que no entrega información valiosa nueva, siendo más bien un contexto de trabajo. En cambio, el ítem 2 ("La pirata Antonia investigó y gritó: ¡La botella 3 es la más valiosa!") explícita una información novedosa que intenciona medir la reacción de activación ante lo declarado.

Tras esta aclaración, el juez aceptó la distinción conceptual, aceptó la fundamentación y, en consecuencia, incrementó su puntuación para el ítem 1, elevando su validez final a 0.89.

En el proceso de validación por jueces expertos, la Juez 5 realizó observaciones críticas sobre un ítem específico, cuestionando inicialmente su ubicación en el instrumento y, principalmente, su pertinencia con la dimensión de curiosidad que se pretende medir. En sus comentarios, señaló que las alternativas del reactivo reflejaban componentes afectivos y emocionales vinculados a la motivación y metacognición, lo que, a su juicio, desvirtuaba su alineación con el constructo. Esto derivó en calificaciones bajas (1) en pertinencia y relevancia. Tras una reunión de clarificación en la que se expuso detalladamente la fundamentación teórica del ítem, se explicó que su ubicación es la segunda posición del instrumento, y que se enfocaba

específicamente en la Curiosidad de Tipo I, característica A4: "buscar nueva información para estimular sentimientos positivos". Se destacó además que la alternativa D representaba de manera clara el deseo de explorar movilizado por emociones positivas, alineándose así con el marco conceptual. Esta justificación permitió a la Juez 5 reevaluar su percepción, lo que se tradujo en un aumento significativo de su puntuación en pertinencia, de 1 a 4 puntos, confirmando la adecuación teórica y operacional del ítem tras la aclaración proporcionada. Dichas puntuaciones se aprecian en la Tabla 12 que nos enseña el incremento de los jueces 4 y 5.

Tabla 12

Resultados finales del grado de acuerdo según el criterio de evaluación de Pertinencia (V de Aiken)

Ítem	Juez 1	Juez 2	Juez 3	Juez 4	Juez 5	Juez 6	V de Aiken
1	4	4	3	4	4	3	0.89
2	4	4	3	4	4	4	0.94
3	4	4	3	4	4	3	0.89
4	4	4	3	4	4	4	0.94
5	4	4	3	4	4	3	0.89
6	4	3	2	4	4	4	0.83
7	4	4	3	4	4	3	0.89

8 4 4 3 4 4 4 0.94

4.2.4 Relevancia

Juzga la contribución esencial de cada ítem a los objetivos generales de la investigación. Un ítem relevante proporciona información única y significativa para la medición del constructo, justificando su inclusión en el instrumento final para evitar redundancias y maximizar la eficiencia de la evaluación. La evaluación inicial del criterio de relevancia se organizó tal como se aprecia en la Tabla 13 para obtener promedios con V de Aiken por cada ítem.

Tabla 13

Resultados iniciales del grado de acuerdo según el criterio de evaluación de Relevancia (V de Aiken)

Ítem	Juez 1	Juez 2	Juez 3	Juez 4	Juez 5	Juez 6	V de Aiken
1	4	2	4	4	4	3	0.89
2	4	4	4	4	4	4	1.00
3	4	4	4	4	4	4	1.00
4	4	4	3	4	1	3	0.72
5	4	4	3	4	4	2	0.83
6	4	4	2	4	4	4	0.89
7	4	4	3	4	4	3	0.89

8 4 4 3 4 4 4 0.94

La aplicación de este criterio arrojó resultados favorables en la mayoría de los ítems, lo que evidencia una alta relevancia general del instrumento de medición. No obstante, se identificó una excepción significativa en el ítem 4, donde el Juez 5 asignó una puntuación de 1 en el criterio de relevancia, un promedio inicial alarmante que requirió atención inmediata. Tras una reunión de clarificación, el juez manifestó que su baja puntuación se debía a una percepción de falta de relevancia y conectividad con la subdimensión objetivo. Este cuestionamiento surgió de una interpretación errónea de la definición de la "curiosidad epistémica tipo I" donde la subdimensión apunta a "Buscar nueva información para estimular el afecto positivo". Una vez realizadas las aclaraciones conceptuales pertinentes, el juez incrementó su puntuación a 4, lo que elevó significativamente el coeficiente V de Aiken del ítem hasta un valor final de 0.89.

Otro ítem que experimentó una mejora en su puntuación tras ajustes menores pero cruciales fue el ítem 5, cuyo valor pasó de 0.83 a 0.89. Esta mejora se logró tras modificar la alternativa de respuesta A. La opción original, "Me gusta ayudarlos, necesito saber qué botella es más valiosa", fue reemplazada por "Me gustaría ayudarlos y me preocupa saber qué botella es más valiosa". Este refinamiento en la redacción fue clave para que el Juez 6 asignara una puntuación más alta, optimizando así la validez del ítem e incrementando el promedio de V de Aiken como lo observamos en la recopilación de datos finales (Tabla 14).

Tabla 14

Resultados finales del grado de acuerdo según el criterio de evaluación de Relevancia (V de Aiken)

Ítem	Juez 1	Juez 2	Juez 3	Juez 4	Juez 5	Juez 6	V de Aiken
1	4	2	4	4	4	3	0.89
2	4	4	4	4	4	4	1.00
3	4	4	4	4	4	4	1.00
4	4	4	3	4	4	3	0.89
5	4	4	3	4	4	3	0.89
6	4	4	2	4	4	4	0.89
7	4	4	3	4	4	3	0.89
8	4	4	3	4	4	4	0.94

4.2.5. Suficiencia

Analiza, a nivel de cada dimensión, si el conjunto de ítems propuestos cubre de manera integral y representativa todas las características definitorias de la misma. Este criterio global garantiza que el conjunto de ítems captura la complejidad y los matices de la curiosidad de interés (tipo I) y por privación (tipo D), sin dejar aspectos clave sin evaluar. Dichos resultados fueron organizados por ítem y puntuación de jueces (Tabla 15)

Tabla 15

Resultados iniciales del grado de acuerdo según el criterio de evaluación de Suficiencia (V de Aiken)

Ítem	Juez 1	Juez 2	Juez 3	Juez 4	Juez 5	Juez 6	V de Aiken
1	4	NA	3	3	NA	NA	0.78
2	4	NA	4	3	4	NA	0.92
3	4	4	3	4	4	NA	0.93
4	4	NA	3	3	NA	NA	0.78
5	NA	NA	3	4	4	NA	0.89
6	NA	NA	2	3	4	NA	0.67
7	NA	NA	3	3	4	NA	0.78
8	NA	NA	4	4	4	NA	1.00

Como se evidencia en la Tabla 15, el criterio de suficiencia registró una elevada frecuencia de "No Aplicable" (NA). Esta situación requirió una investigación inmediata para determinar su origen, para lo cual se contactó a los jueces 1, 2, 5 y 6. Tras la consulta, se identificó que la causa principal era una incomprensión de lo que se debía evaluar en los ítems específicos pues todos los jueces manifestaron esta inquietud. En consecuencia, se procedió a explicar el objetivo del criterio de suficiencia, detallado al inicio del apartado. Tras esta aclaración, los jueces asignaron nuevas puntuaciones, predominantemente 3 y 4 (ver Tabla 16), lo que indica una alta pertinencia del instrumento. No obstante, para futuras implementaciones, se recomienda incluir al menos dos ítems por característica de cada tipo de curiosidad epistémica (por privación y por interés) para garantizar una mayor cobertura y una evaluación más robusta de la suficiencia.

Tabla 16

Resultados finales del grado de acuerdo según el criterio de evaluación de Suficiencia (V de Aiken)

Ítem	Juez 1	Juez 2	Juez 3	Juez 4	Juez 5	Juez 6	V de Aiken
1	4	4	3	3	4	3	0.83
2	4	4	4	3	4	3	0.89
3	4	4	3	4	4	3	0.89
4	4	4	3	3	4	3	0.83
5	4	4	3	4	4	3	0.89
6	4	4	2	3	4	3	0.78
7	4	4	3	3	4	3	0.83
8	4	4	4	4	4	3	0.94

4.2.6 Evaluación final de la validez de contenido tras la revisión por jueces expertos

Tras el proceso de revisión y ajuste de los ítems, se llevó a cabo una segunda ronda de evaluación con el mismo panel de expertos con el objetivo de determinar el nivel de validez de contenido de la versión final del instrumento. Los resultados de esta evaluación que se presentan en la Tabla 9, permiten constatar el impacto positivo de las modificaciones realizadas. Para un análisis integral, se ha calculado el promedio total del V de Aiken por ítem, el cual ofrece una visión global de su aceptación final Tabla 17.

Tabla 17

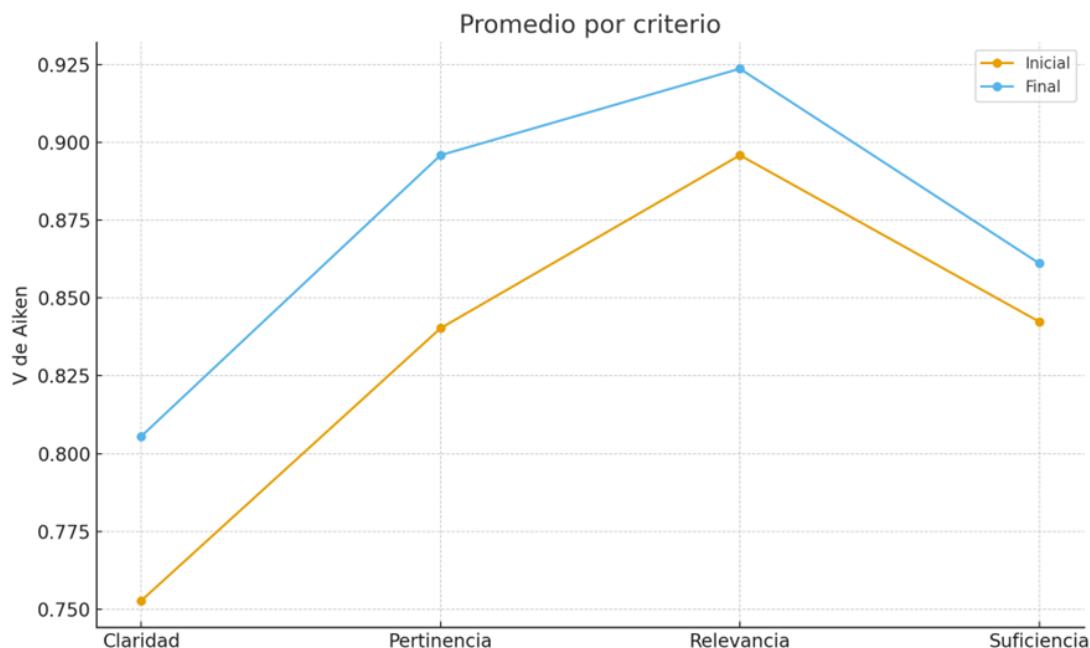
Valores finales de V de Aiken por ítem y criterio y promedio por ítem.

Ítem	Claridad	Pertinencia	Relevancia	Suficiencia	Promedio (V)
1	0.889	0.889	0.889	0.833	0.875
2	0.889	0.944	1.000	0.889	0.931
3	0.778	0.889	1.000	0.889	0.889
4	0.833	0.944	0.889	0.833	0.875
5	0.778	0.889	0.889	0.889	0.861
6	0.778	0.833	0.889	0.778	0.820
7	0.722	0.889	0.889	0.833	0.833
8	0.778	0.889	0.944	0.944	0.889

A continuación, en la Figura 8 se presentan de forma detallada los resultados cuantitativos y gráficos de ambas rondas de evaluación para facilitar la comprensión de los resultados de validación de jueces expertos.

Figura 8

Promedios iniciales y finales por criterio



Para la interpretación de los resultados finales, se mantuvieron los mismos rangos de clasificación del coeficiente V de Aiken utilizados en la evaluación inicial (Aiken, 1985; Penfield y Giacobbi, 2004), garantizando así la consistencia metodológica en el proceso de validación. La aplicación de estos criterios a los promedios finales revela una mejora sustancial en la calidad global del instrumento. En la categoría de Validez Excelente ($V \geq 0.90$), los ítems 2 ($V=0.931$) y 8 ($V=0.889$) no solo mantuvieron, sino que, en el caso del ítem 2, consolidaron su estatus de excelencia. A este grupo se une el ítem 1 ($V=0.875$), el cual, tras las revisiones realizadas, alcanzó este nivel de excelencia, confirmando que estos ítems son óptimos y no requieren modificaciones adicionales. Por otro lado, en la categoría de validez buena a muy buena ($0.80 \leq V < 0.90$), se concentra la mayoría de los ítems, demostrando un nivel de calidad alto. Los ítems 3 ($V=0.889$), 4 ($V=0.875$), 5 ($V=0.861$), 7 ($V=0.833$) y 8 ($V=0.889$) presentan una validez muy buena,

destacándose el progreso significativo del ítem 3, que mejoró desde una categoría de validez aceptable/baja en la ronda inicial. Asimismo, el ítem 6 ($V=0.820$) se sitúa en este nivel, lo que indica que las revisiones realizadas fueron efectivas para elevarlo por encima del umbral de 0.80.

En conclusión, el proceso de validación de contenido en dos rondas demostró ser altamente efectivo. La primera ronda sirvió como un diagnóstico preciso que identificó las debilidades específicas del instrumento inicial, mientras que la fase de reformulación, guiada por la retroalimentación cualitativa de los expertos, permitió abordar de manera dirigida problemas de ambigüedad, solapamiento semántico e insuficiencia informativa. Los resultados de la segunda ronda, sintetizados en la Tabla 20, evidencian el éxito de este proceso iterativo, al observarse una mejora generalizada y consistente en los valores del V de Aiken. La totalidad de los ítems alcanzó o superó el umbral de 0.80 en su promedio final, ubicándose en las categorías de validez buena a muy buena y excelente. Este hecho corrobora que el instrumento final posee unos estándares de claridad, pertinencia, relevancia y suficiencia óptimos, garantizando su validez de contenido para su aplicación en el contexto de estudio previsto. Por lo tanto, el instrumento se considera válido y robusto para proceder a las siguientes fases de la investigación.

4.3 RESULTADOS DE APLICACIÓN PILOTO

La evaluación psicométrica de instrumentos educativos requiere contrastar los resultados empíricos con estándares técnicos establecidos para garantizar validez y confiabilidad. Siguiendo los criterios clásicos de la teoría de respuesta al ítem y la teoría clásica de tests, un ítem se considera con dificultad adecuada cuando su índice p se sitúa entre 0.30 y 0.70 en poblaciones homogéneas (Haladyna, 2004), mientras que valores por debajo de 0.20 indican una dificultad extrema y posibles problemas de constructo. En cuanto a la discriminación, una correlación ítem-total

corregida superior a 0.30 se considera aceptable (Nunnally y Bernstein, 1994), y valores negativos o cercanos a cero revelan que el ítem no diferencia entre estudiantes con alto y bajo rasgo latente. Finalmente, para la confiabilidad interna, un alfa de Cronbach ≥ 0.70 es el mínimo recomendado para instrumentos de investigación, mientras que valores inferiores a 0.60 se consideran inaceptables para la toma de decisiones pedagógicas (Cortina, 1993). Se decidió realizar la prueba en estudiantes sin necesidades educativas especiales.

4.3.1 Análisis de los resultados de la aplicación piloto para segundo básico

Los resultados de la aplicación piloto del instrumento de curiosidad epistémica en 21 estudiantes de segundo básico muestran un desempeño psicométrico deficiente en múltiples dimensiones. Como se presenta en la Tabla 18, el puntaje total presentó una media baja ($M=2.81$, $DE=1.29$), con un rango observado de 0 a 6 sobre un máximo posible de 8, y una distribución donde el 66% de los estudiantes se concentró entre 2 y 4 puntos. Esto sugiere una limitada variabilidad y un posible techo de dificultad.

Tabla 18

Estadísticos descriptivos del puntaje total (N=21)

Estadístico	Valor
Media	2.81
Desviación Estándar	1.29

Mínimo 0.00

Máximo 6.00

Al analizar las dimensiones de Curiosidad de Interés (A) y Curiosidad por Privación (B) mediante el índice de dificultad (p) por ítem (Tabla 19), se observa que los ítems asociados a las subdimensiones A2 (activación ante lo nuevo, ítem 1: $p=.20$), A3 (disfrute de descubrimientos, ítem 3: $p=.43$), B2 (estimulación por carencia, ítem 4: $p=.40$), B3 (ansiedad por incertidumbre, ítem 7: $p=.14$) y B4 (reducción de incógnitas, ítem 8: $p=.095$) presentan índices de dificultad críticos. En particular, los ítems B3 y B4 se sitúan muy por debajo del estándar mínimo de $p=.20$, lo que los clasifica como extremadamente difíciles y posiblemente mal calibrados para el nivel cognitivo del grupo evaluado.

Tabla 19

Índice de dificultad (p) por ítem

Ítem	Dimensión	p	Clasificación
1	A2	.20	Muy difícil
2	A4	.52	Aceptable
3	A3	.43	Aceptable

4	B2	.40	Aceptable-difícil
5	A1	.62	Aceptable-fácil
6	B1	.43	Aceptable
7	B3	.14	Extremadamente difícil
8	B4	.095	Extremadamente difícil

La distribución de alternativas por ítem (Tabla 20) revela un comportamiento problemático de los distractores, con distribuciones uniformes en varios ítems (ej. ítem 1: alternativa 3 seleccionada por el 47.6%) y una captura casi total de la muestra por distractores en los ítems más difíciles (B3 y B4), lo que sugiere ambigüedad en la redacción o un desajuste con el nivel lector.

Tabla 20

Distribución de alternativas por ítem (resumen)

Ítem	Alternativa más seleccionada	%	Dimensión asociada
1	3	47.6%	A2

2	4	52.4%	A4
3	3	42.9%	A3
4	3	38.1%	B2
5	1	61.9%	A1
6	1	42.9%	B1
7	2	57.1%	B3
8	1	57.1%	B4

La capacidad discriminativa del instrumento también resultó inadecuada. Según la matriz de correlaciones ítem-total corregida (Tabla 21), varios ítems mostraron valores negativos o cercanos a cero, destacándose los ítems de las dimensiones B3 ($r=-.121$) y B4 ($r=.022$), los cuales no cumplen con el criterio mínimo de $r>.30$. Esto indica que estos reactivos no logran diferenciar entre estudiantes con mayor o menor curiosidad epistémica, e incluso podrían estar penalizando a aquellos con mayor rendimiento global.

Tabla 21*Correlaciones ítem-total corregida (resumen)*

Ítem	Dimensión	Correlación ítem-total corregida	Clasificación
1	A2	.356	Moderada
2	A4	.110	Débil
3	A3	.560	Aceptable
4	B2	.461	Moderada
5	A1	.318	Aceptable-límite
6	B1	.461	Moderada
7	B3	-.121	Negativa (inaceptable)
8	B4	.022	Nula (inaceptable)

Finalmente, la confiabilidad interna medida mediante el alfa de Cronbach fue de $\alpha=.502$ (Tabla 22), valor muy por debajo del estándar aceptable de .70, y acompañado de una advertencia

del sistema sobre covarianzas negativas entre ítems, lo que sugiere una falta de homogeneidad en la medición del constructo (Cronbach, 1951).

Tabla 22

Estadísticas de fiabilidad

Estadístico	Valor	Interpretación
Alfa de Cronbach	.502	Inaceptable
Alfa basada en elementos estandarizados	.262	Crítico
N de elementos	9	-

La síntesis de los problemas anteriores se manifiesta en el coeficiente de confiabilidad. El alfa de Cronbach obtenido fue de $\alpha = .502$ (Tabla 22), muy por debajo del mínimo aceptable de .70 para investigación (Cortina, 1993). Es crucial destacar que el valor del "alfa basado en elementos estandarizados" fue aún más bajo (.262), lo que indica que ni siquiera la estandarización de las respuestas corrige la falta de covariación entre los ítems. El sistema estadístico emitió una advertencia sobre covarianzas negativas en la matriz, confirmada por la correlación negativa del ítem 7. Este resultado no es simplemente "baja confiabilidad"; es evidencia de que el conjunto de ítems no está midiendo un constructo unidimensional de manera coherente. Algunos ítems (ej., 3, 4, 6) apuntan en una dirección, mientras que otros (7 y, en menor medida, 2) apuntan en otra o en ninguna.

Los resultados de la aplicación piloto no invalidan el constructo teórico de la curiosidad epistémica, pero sí demuestran de manera concluyente que su operacionalización en el instrumento actual es psicométricamente inviable y evolutivamente inadecuada para estudiantes de segundo año básico. El análisis integrado sugiere lo siguiente:

1. Problema de validez de contenido: Las dimensiones más complejas y abstractas del modelo (particularmente las de la curiosidad por privación - B3 y B4) no son susceptibles de ser evaluadas mediante auto-reporte en esta edad. Los ítems que las miden son extremadamente difíciles, no discriminan y deterioran la consistencia interna.
2. Problema de formulación lingüística: El funcionamiento anómalo de los distractores y la baja dificultad de ítems incluso de dimensiones "asequibles" (como A2) indican que el lenguaje, la extensión de las frases o el vocabulario utilizado no son apropiados para el nivel lector y comprensivo del grupo.
3. Problema de método de respuesta: La concentración de respuestas en pocas alternativas y la baja varianza sugieren que el formato de respuesta (ej., escala Likert con muchas opciones) puede ser confuso. Para esta edad, se recomiendan formatos más simples (ej., caras con expresiones, opciones binarias ilustradas).

Directrices concretas para la revisión y reestructuración de los ítems

- Eliminación: Se recomienda la eliminación inmediata de los ítems 7 y 8, dados sus índices psicométricos catastróficos y su probable inadecuación teórica para la edad.
- Reformulación profunda: Los ítems 1, 2 y 4 deben ser redactados nuevamente, utilizando un vocabulario más concreto, frases más cortas y referentes a situaciones específicas y familiares del aula (ej., "cuando el profesor muestra un objeto nuevo, yo...").
- Revisión de distractores: Todos los distractores deben ser evaluados por juicios expertos (docentes de primer ciclo) para asegurar su plausibilidad, pero no su atractivo por encima de la clave.

- Reconsideración del modelo: Para esta cohorte de edad, es posible que el instrumento deba centrarse principalmente en los aspectos de Curiosidad de Interés (disfrute, reconocimiento, persistencia), que mostraron un funcionamiento más promisorio, y postergar o reformular radicalmente los componentes de "privación" para edades mayores.
- Nueva prueba piloto: El instrumento revisado debe someterse a una nueva prueba piloto, idealmente complementada con entrevistas cognitivas a una submuestra de niños, para comprender su proceso de interpretación y respuesta.

Esta etapa piloto, por tanto, cumple su objetivo esencial que es detectar fallas críticas antes de la aplicación principal y proporcionar un diagnóstico fundamentado para una revisión sistemática que incremente las probabilidades de éxito en la medición válida y confiable del constructo en esta población específica.

4.3.2 Análisis de los resultados de la aplicación piloto en tercero básico

La evaluación psicométrica del instrumento aplicado a 24 estudiantes de tercero básico sigue los mismos estándares técnicos previamente establecidos: un índice de dificultad (p) entre 0.30 y 0.70 se considera adecuado (Haladyna, 2004), una correlación ítem-total corregida mayor a 0.30 indica una discriminación aceptable (Nunnally y Bernstein, 1994), y un alfa de Cronbach \geq 0.70 es el mínimo requerido para garantizar una confiabilidad interna aceptable (Cortina, 1993).

Los resultados descriptivos del puntaje total (Tabla 23) muestran un desempeño general moderado, con una distribución que va de 0 a 6 puntos (sobre 8 posibles). La concentración de respuestas en los puntajes 4 y 5 (50% de la muestra) sugiere una dificultad global adecuada, aunque

= .542) y B1 (placer intrínseco y preocupaciones extrínsecas, ítem 6: $p = .565$) se encuentran dentro del rango de dificultad aceptable. Sin embargo, los ítems B3 (ansiedad por incertidumbre, ítem 7: $p = .125$) y B4 (reducción de incógnitas, ítem 8: $p = .09$) presentan una dificultad extrema, muy por debajo del estándar mínimo ($p < .20$), lo que indica un grave desajuste con las capacidades del grupo evaluado.

Tabla 24

Índice de dificultad (p) por ítem y dimensión asociada

Ítem	Dimensión	p	Clasificación
1	A2	0.50	Adecuado
2	A4	0.667	Aceptable-fácil
3	A3	0.50	Adecuado
4	B2	0.583	Aceptable
5	A1	0.542	Aceptable
6	B1	0.565	Aceptable

7	B3	0.125	Extremadamente difícil
8	B4	0.09	Extremadamente difícil

La distribución de alternativas por ítem (Tabla 25) revela un comportamiento diferenciado de los distractores. Para los ítems 1 al 6, la alternativa correcta suele ser la más seleccionada, lo que sugiere un funcionamiento aceptable. No obstante, en los ítems 7 (B3) y 8 (B4), los distractores capturan entre el 87.5% y el 91% de las respuestas, respectivamente, indicando que son demasiado fuertes, plausibles o que la clave correcta es ambigua o inaccesible para los estudiantes.

Tabla 25

Distribución de alternativas por ítem (resumen)

Ítem	Alternativa correcta	% de acierto	Dimensión	Comportamiento del distractor
1	1	50.0%	A2	Funcional
2	4	66.7%	A4	Funcional
3	3	50.0%	A3	Funcional
4	3	58.3%	B2	Funcional

5	1	54.2%	A1	Funcional
6	1	56.5%	B1	Funcional
7	1	12.5%	B3	Demasiado fuertes
8	1	9.1%	B4	Demasiado fuertes

La capacidad discriminativa del instrumento, medida mediante la correlación ítem-total corregida (Tabla 26), es deficiente. Ningún ítem alcanza el valor de referencia de 0.30. El ítem 4 (B2) presenta la correlación más alta (0.259), aún insuficiente. El ítem 8 (B4) muestra una correlación negativa (-0.046), lo que es técnicamente inaceptable, ya que indica que los estudiantes con mayor puntaje total tienden a fallar ese ítem, revelando un grave problema de diseño o validez de constructo.

Tabla 26

Correlación ítem-total corregida y estadísticas de fiabilidad por ítem

Ítem	Dimensión	Correlación Ítem-Total	Alfa si se elimina
------	-----------	------------------------	--------------------

1	A2	0.187	0.296
2	A4	0.197	0.293
3	A3	0.150	0.318
4	B2	0.259	0.254
5	A1	0.107	0.342
6	B1	0.150	0.318
7	B3	0.093	0.343
8	B4	-0.046	0.387

Finalmente, la confiabilidad interna del instrumento (Tabla 27) es muy baja, con un alfa de Cronbach de 0.351 y un alfa estandarizado de 0.327, valores muy por debajo del mínimo aceptable de 0.70. Esto confirma que los ítems no miden de manera consistente un constructo único y homogéneo de curiosidad epistémica, invalidando el uso del puntaje total como un indicador fiable.

Tabla 27*Estadísticas de fiabilidad del instrumento*

Estadístico	Valor	Interpretación
Alfa de Cronbach	0.351	Inaceptable
Alfa basado en elementos estandarizados	0.327	Crítico
N de elementos	8	-

En síntesis, podemos evidenciar que el instrumento aplicado en tercero básico muestra una dificultad general más equilibrada que en segundo básico especialmente en las dimensiones de Curiosidad de Interés (A), persisten problemas psicométricos severos. La discriminación es deficiente en todos los ítems, la confiabilidad interna es insuficiente y los ítems de las dimensiones de Privación (B3 y B4) son extremadamente difíciles y presentan distractores disfuncionales. Estos hallazgos respaldan la necesidad de una revisión integral del instrumento, con especial atención a la reformulación de los ítems de la dimensión de privación, el rediseño de distractores y una nueva validación de contenido acorde al nivel de desarrollo de los estudiantes de tercer grado.

La evaluación psicométrica del instrumento aplicado a 24 estudiantes de tercero básico confirma problemas estructurales severos, aunque se observa una ligera mejoría en la dificultad de

los ítems en comparación con el curso analizado anteriormente. El estudio revela una incoherencia interna crítica del instrumento.

Evidencia de multidimensionalidad no deseada, el coeficiente de confiabilidad es insatisfactorio para el estudio realizado, con un alfa de Cronbach de $\alpha = .351$ y un alfa estandarizado de .327, muy por debajo del mínimo de .70. Este resultado no indica simplemente baja confiabilidad, sino que los ítems no covarían de manera consistente para medir un constructo unificado. La correlación ítem-total corregida es deficiente en todos los casos (ningún ítem supera el umbral de .30), y el ítem 8 (B4) incluso presenta una correlación negativa (-.046), señalando que los estudiantes con mayor puntaje total tienden un fallo psicométrico grave.

Dificultad y funcionamiento de ítems: Un patrón divergente. Se identifica un claro patrón según la dimensión teórica:

- Dimensiones de Curiosidad de Interés (A) y de Privación (B): Los ítems 1 (A2), 2 (A4), 3 (A3), 5 (A1) y 6 (B1) muestran un índice de dificultad (p) dentro del rango aceptable (.50 a .667) y sus distractores funcionan adecuadamente, siendo la clave la alternativa más seleccionada.
- Dimensiones de Curiosidad por Privación (B - componentes complejos): Los ítems 7 (B3, $p = .125$) y 8 (B4, $p = .09$) son extremadamente difíciles. Sus distractores capturan entre el 87.5% y el 91% de las respuestas, lo que demuestra que son demasiado fuertes o que los conceptos subyacentes (ansiedad por incertidumbre, reducción de incógnitas) son inaccesibles para la gran mayoría de los estudiantes de este nivel.

Si bien el instrumento presenta una dificultad global más equilibrada que en segundo básico, especialmente para las dimensiones de "Interés", psicométricamente no es viable. La falta

de discriminación de todos los ítems y la confiabilidad interna nula, unidas a la extrema dificultad de los ítems de "Privación", invalidan el uso del puntaje total. Los resultados refuerzan la necesidad de una revisión profunda, particularmente en la operacionalización de los componentes más abstractos del constructo.

4.3.3 Análisis de los resultados de la aplicación piloto en cuarto básico

La evaluación psicométrica del instrumento aplicado a 30 estudiantes de cuarto básico mantiene como referentes los estándares establecidos por la teoría clásica de tests: un índice de dificultad (p) entre 0.30 y 0.70 se considera óptimo para la mayoría de los contextos (Haladyna, 2004), una correlación ítem-total corregida mayor a 0.30 indica una discriminación adecuada (Nunnally y Bernstein, 1994), y un alfa de Cronbach ≥ 0.70 representa el mínimo de confiabilidad interna necesario para interpretaciones grupales (Cortina, 1993). El análisis de los resultados de este nivel revela una mejora sustancial en comparación con los grados anteriores, aunque persisten desafíos críticos.

El desempeño general del grupo se presenta en la Tabla 28, que muestra una distribución del puntaje total (corregido) que oscila entre 1 y 6 puntos, con una concentración en los puntajes medios-altos. El 30% de los estudiantes obtuvo 5 puntos y el 23.3% obtuvo 3 puntos. Si bien ningún estudiante alcanzó el puntaje máximo de 8, la distribución sugiere una mayor adecuación del instrumento al nivel cognitivo de cuarto básico, con una dispersión más saludable de los resultados.

Tabla 28

Distribución del puntaje total corregido (N=30)

Puntaje	Frecuencia	Porcentaje
1	1	3.3%
2	4	13.3%
3	7	23.3%
4	6	20.0%
5	9	30.0%
6	3	10.0%

El análisis del índice de dificultad (p) por ítem (Tabla 29) muestra un panorama mixto y alineado con las dimensiones teóricas. Los ítems correspondientes a las dimensiones de Curiosidad de Interés presentan un comportamiento generalmente favorable: A2 (activación ante lo nuevo, ítem 1: $p = .43$) y A3 (disfrute de descubrimientos, ítem 3: $p = .59$) están dentro del rango óptimo. Sin embargo, A4 (búsqueda de información para estímulos positivos, ítem 2: $p = .87$) y A1 (motivación intrínseca, ítem 5: $p = .83$) son excesivamente fáciles, lo que sugiere que los estudiantes comprenden fácilmente estos conceptos o que los ítems carecen de desafío. Para las dimensiones de Curiosidad por Privación, los resultados son más problemáticos: B2 (estimulación por carencia, ítem 4: $p = .70$) es aceptable aunque en el límite superior de facilidad, B1 (placer

intrínseco y preocupaciones extrínsecas, ítem 6: $p = .33$) es difícil pero dentro del rango funcional, mientras que B3 (ansiedad por incertidumbre, ítem 7: $p = .10$) y B4 (reducción de incógnitas, ítem 8: $p = .07$) son extremadamente difíciles, replicando el patrón crítico observado en los cursos anteriores.

Tabla 29

Índice de dificultad (p) por ítem y dimensión asociada

Ítem	Dimensión	p	Clasificación
1	A2	0.433	Adecuado
2	A4	0.867	Muy fácil
3	A3	0.586	Aceptable
4	B2	0.724	Aceptable-fácil
5	A1	0.833	Muy fácil
6	B1	0.333	Difícil, pero aceptable
7	B3	0.10	Extremadamente difícil
8	B4	0.07	Extremadamente difícil

La distribución de alternativas por ítem (Tabla 30) corrobora los hallazgos de dificultad. En los ítems 1, 3, 4 y 6, la alternativa correcta es la más seleccionada, indicando un funcionamiento adecuado de los distractores. En contraste, en los ítems 7 y 8 los distractores son abrumadoramente dominantes (ej. en ítem 7, un distractor atrae al 70% de las respuestas), lo que evidencia una grave falla en su diseño, ya sea por una clave poco clara o por distractores demasiado plausibles para el nivel de comprensión de los estudiantes.

Tabla 30

Distribución de alternativas por ítem (resumen)

Ítem	Alternativa correcta	% de acierto	Dimensión	Comportamiento del distractor
1	1	43.3%	A2	Funcional
2	4	86.7%	A4	Muy fácil, distractor débil
3	3	58.6%	A3	Funcional
4	3	72.4%	B2	Funcional
5	1	83.3%	A1	Muy fácil, distractor débil
6	4	33.3%	B1	Funcional

7	1	10.0%	B3	Distractores dominantes
8	1	7.1%	B4	Distractores dominantes

La capacidad discriminativa de los ítems, evaluada mediante la correlación ítem-total corregida (Tabla 31), presenta resultados muy variados. El ítem 1 (A2) destaca con una discriminación excelente (0.694). Los ítems 4 (B2: 0.436) y 5 (A1: 0.376) muestran una discriminación moderada y aceptable. Sin embargo, los ítems 2 (A4: 0.122), 3 (A3: 0.157), 6 (B1: 0.223) y 7 (B3: 0.198) presentan una discriminación baja, lo que limita su utilidad para diferenciar entre niveles de curiosidad. El ítem 8 (B4) nuevamente muestra un valor preocupante (-0.043), indicando una discriminación negativa y un funcionamiento psicométricamente inverso.

Tabla 31

Correlación ítem-total corregida y efecto en la confiabilidad

Ítem	Dimensión	Correlación Ítem-Total	Alfa si se elimina
1	A2	0.694	0.543
2	A4	0.122	0.650

3	A3	0.157	0.649
4	B2	0.436	0.598
5	A1	0.376	0.613
6	B1	0.223	0.636
7	B3	0.198	0.641
8	B4	-0.043	0.660

Finalmente, la confiabilidad interna del instrumento (Tabla 32), si bien muestra la mejora más significativa en comparación con los niveles anteriores, aún no alcanza el estándar mínimo. El alfa de Cronbach es de 0.644 y el alfa estandarizado es de 0.547. Este valor, aunque cercano al umbral de 0.70, sigue siendo insuficiente para garantizar que el puntaje total sea una medida estable y consistente del constructo de curiosidad epistémica en este nivel.

Tabla 32

Estadísticas de fiabilidad del instrumento

Estadístico	Valor	Interpretación
-------------	-------	----------------

Alfa de Cronbach	0.644	Insuficiente (cercano al límite)
Alfa basado en elementos estandarizados	0.547	Bajo
N de elementos	8	-

En síntesis, el análisis para cuarto básico revela un instrumento en transición. Por un lado, se observa una clara mejoría psicométrica: la dificultad de la mayoría de los ítems se sitúa en rangos aceptables, y uno de ellos (A2) presenta una discriminación sobresaliente. Esto sugiere que las dimensiones de Curiosidad de Interés, en particular, son más accesibles y medibles en este nivel. No obstante, los problemas estructurales persisten: la dimensión de Curiosidad por Privación sigue siendo extremadamente difícil de operacionalizar (ítems B3 y B4), la confiabilidad total es aún inaceptable, y la presencia de ítems demasiado fáciles (A4, A1) y con discriminación negativa (B4) compromete la calidad métrica global. Estos hallazgos confirman que, si bien el instrumento es más viable en cuarto básico que en cursos inferiores, requiere una revisión focalizada especialmente en los ítems de privación y en aquellos de baja discriminación antes de poder ser considerado válido y confiable para su uso evaluativo.

La evaluación en 30 estudiantes de cuarto básico muestra la evolución más prometedora, con una clara mejoría métrica que sugiere una mayor adecuación del instrumento al desarrollo cognitivo de este nivel. Sin embargo, persisten desafíos fundamentales que impiden su aplicación confiable.

Mejoría insuficiente. El alfa de Cronbach ($\alpha = .644$) y el alfa estandarizado (.547) representan una mejora sustancial respecto a cursos anteriores, pero siguen por debajo del estándar mínimo aceptable (.70). Esto indica que, si bien los ítems comienzan a covariar mejor, la medida global aún carece de la consistencia interna necesaria para interpretaciones grupales válidas.

Patrón consolidado y aparición de nuevos problemas. El análisis por dimensiones confirma y matiza los hallazgos previos:

- Curiosidad de Interés (A): Los ítems 1 (A2, $p=.43$) y 3 (A3, $p=.59$) están en el rango óptimo. No obstante, los ítems 2 (A4, $p=.87$) y 5 (A1, $p=.83$) son excesivamente fáciles, sugiriendo un techo que limita su capacidad para discriminar.
- Curiosidad por Privación (B): Se replica el patrón crítico: los ítems 7 (B3, $p=.10$) y 8 (B4, $p=.07$) son extremos en su dificultad. Los distractores dominan las respuestas (hasta un 90%), evidenciando la incomprensión o inadecuación de estos ítems. Los ítems 4 (B2, $p=.72$) y 6 (B1, $p=.33$) muestran un funcionamiento aceptable.

Resultados heterogéneos y una señal de alarma. La capacidad de los ítems para diferenciar entre estudiantes con mayor y menor curiosidad es desigual. El ítem 1 (A2) tiene una discriminación excelente (.694), y los ítems 4 (B2, .436) y 5 (A1, .376) son aceptables. Sin embargo, los ítems 2, 3, 6 y 7 discriminan poco ($<.30$). Críticamente, el ítem 8 (B4) nuevamente muestra una correlación ítem-total negativa (-.043), un defecto psicométrico inaceptable que confirma su mal funcionamiento sistemático.

El instrumento muestra una evolución positiva pero incompleta. Las dimensiones de "Interés" se vuelven más medibles, con un ítem (A2) de funcionamiento óptimo. Sin embargo, los problemas estructurales son definitivos: la confiabilidad total sigue siendo insuficiente, la

dimensión de "Privación" (especialmente B3 y B4) es psicométricamente inviable, y la aparición de ítems demasiado fáciles y con discriminación negativa exige una revisión profunda. Estos hallazgos indican que el instrumento actual, aunque más cercano a la viabilidad en cuarto básico, requiere una reestructuración focalizada antes de cualquier aplicación con fines evaluativos.

4.3.4 Análisis final de los resultados de la aplicación piloto

La aplicación piloto del instrumento de curiosidad epistémica en estudiantes de segundo, tercero y cuarto año básico ha proporcionado evidencia empírica decisiva sobre su viabilidad psicométrica y su adecuación al desarrollo cognitivo de la población objetivo. Los resultados, analizados bajo los estándares de la teoría clásica de tests y la teoría de respuesta al ítem, permiten establecer una conclusión clara y fundamentada:

El instrumento en su forma actual no es válido ni confiable para medir el constructo de curiosidad epistémica en los niveles evaluados. Si bien se observa una mejora progresiva en los indicadores psicométricos conforme aumenta el grado escolar especialmente en cuarto básico, ninguno de los tres grupos alcanzó los estándares mínimos de calidad métrica requeridos para instrumentos de investigación. La confiabilidad interna fue consistentemente inaceptable ($\alpha = .502$, $.351$ y $.644$ para segundo, tercero y cuarto básico, respectivamente), con valores de alfa estandarizado aún más bajos y advertencias de covarianzas negativas entre ítems. Esto indica que el conjunto de reactivos no mide un constructo unidimensional de manera coherente.

Se identificó un patrón crítico diferenciado por dimensiones teóricas: los ítems correspondientes a la Curiosidad de Interés (dimensiones A) mostraron un funcionamiento más promisorio, con índices de dificultad generalmente adecuados y, en algunos casos, una discriminación aceptable. En contraste, los ítems de la Curiosidad por Privación (especialmente

las subdimensiones B3 "ansiedad por incertidumbre" y B4 "reducción de incógnitas") presentaron un desempeño psicométricamente catastrófico en los tres niveles: índices de dificultad extremos ($p < .20$), discriminación nula o negativa, y distractores que capturaron de manera abrumadora las respuestas de los estudiantes. Esto sugiere una inadecuación evolutiva y operacional de estos componentes más abstractos para niños de 6 a 10 años, quienes probablemente no pueden autorreportar de manera fiable estos estados cognitivo-afectivos complejos.

Además, se detectaron problemas metodológicos transversales: formulación lingüística inapropiada para el nivel lector, diseño de distractores disfuncionales y un posible formato de respuesta (escala Likert) inadecuado para las edades más tempranas. La baja varianza en las puntuaciones totales y la concentración en alternativas específicas refuerzan la necesidad de adaptar el instrumento al contexto de desarrollo de los estudiantes.

En consecuencia, esta etapa piloto cumple su propósito fundamental: detectar fallas estructurales antes de una aplicación a gran escala y proporcionar un diagnóstico preciso para una revisión sistemática. Las directrices derivadas son concretas: 1) eliminación inmediata de los ítems B3 y B4; 2) reformulación profunda de los ítems restantes con lenguaje concreto, situaciones familiares y frases breves; 3) rediseño completo de los distractores con validación por juicio experto docente; 4) reconsideración del modelo teórico aplicable a esta edad, priorizando la medición de la Curiosidad de Interés; y 5) implementación de una nueva prueba piloto con entrevistas cognitivas que permitan comprender los procesos de interpretación infantil.

Por lo tanto, se concluye que el instrumento requiere una reestructuración integral y focalizada que considere no sólo los criterios psicométricos, sino también los aspectos evolutivos, lingüísticos y contextuales de la población. Solo mediante una revisión fundamentada en estos

hallazgos podrá desarrollarse una herramienta válida y confiable para medir la curiosidad epistémica en el ciclo básico inicial.

V. DISCUSIONES

El presente estudio tuvo como objetivo central diseñar y validar una escala para medir la curiosidad epistémica en contextos de razonamiento inferencial informal en matemáticas, dirigida a estudiantes de segundo a cuarto básico del sistema escolar chileno. La discusión de los resultados se centrará en tres ejes fundamentales: 1) la importancia y aporte del instrumento desarrollado frente al vacío identificado en la literatura; 2) el análisis de la validez de contenido obtenida mediante el juicio de expertos, con énfasis en los valores de la V de Aiken; y 3) la evaluación de los resultados de la aplicación piloto y su relación con las problemáticas y limitaciones planteadas inicialmente.

5.1 Importancia y aporte del instrumento en el contexto nacional e internacional

La revisión sistemática realizada al inicio de esta investigación evidenció una carencia crítica de instrumentos específicos, válidos y confiables para medir la curiosidad epistémica, particularmente en el dominio matemático y en población infantil (Jirout, 2020; McNary, 2023). Estudios como el de Ahmad y Siew (2021) o Tang et al. (2024) han desarrollado escalas en otros contextos, pero presentan limitaciones psicométricas o no se focalizan en el razonamiento matemático. En el ámbito hispanohablante, la validación de la escala de Litman por Sánchez et al. (2023) constituye un avance, pero su diseño para adultos y sus limitaciones muestrales la hacen inaplicable para estudiantes de primaria (Bengart et al., 2018). Este vacío genera un triple problema: para la investigación, imposibilita establecer relaciones precisas entre la curiosidad epistémica y variables clave como el rendimiento académico; para la práctica pedagógica, limita el diseño de intervenciones diferenciadas; y para la política educativa, dificulta evaluar el impacto de las reformas curriculares que promueven la indagación y la curiosidad (MINEDUC, 2018, 2024).

Frente a esta brecha, el instrumento desarrollado en este estudio representa un aporte significativo y original. En primer lugar, opera una doble contextualización: ancla el constructo de curiosidad epistémica en un dominio de conocimiento específico (matemáticas) y lo sitúa en un tipo de razonamiento particular (inferencial informal), alineado con los Objetivos de Aprendizaje del currículum nacional chileno (Estrella, 2025; Ministerio de Educación de Chile, 2018). Esto permite capturar manifestaciones situadas y ecológicamente válidas de la curiosidad, como el interés por descifrar patrones (Tipo-I) o la incomodidad ante una conjetura no resuelta (Tipo-D), en escenarios que son familiares y relevantes para los estudiantes. En segundo lugar, se fundamenta en el modelo bidimensional de Litman (2008), que distingue entre la curiosidad motivada por el interés y el placer de aprender (Tipo-I) y aquella motivada por la privación y la necesidad de reducir la incertidumbre (Tipo-D). Esta distinción teórica, poco explorada en instrumentos para niños, permite un diagnóstico más fino de los perfiles motivacionales de los estudiantes, lo que puede informar prácticas pedagógicas más efectivas.

Por lo tanto, este estudio no solo responde a una necesidad metodológica identificada en la literatura, sino que también proporciona una herramienta con potencial para impulsar la investigación aplicada en educación matemática en Chile, ofreciendo a docentes e investigadores un medio para evaluar un constructo que es reconocido como un motor esencial del aprendizaje profundo y del engagement académico (Jirout, 2020; von Stumm et al., 2011).

5.2 Análisis de la validez de contenido: El rol de la V de Aiken y el juicio de expertos

La validez de contenido es un requisito fundamental en el desarrollo de instrumentos, ya que garantiza que los ítems representen adecuadamente el constructo teórico que se pretende medir (American Educational Research Association et al., 2014; Haynes et al., 1995). En este estudio, se

adoptó un procedimiento riguroso de validación por juicio de expertos, utilizando el coeficiente V de Aiken como medida cuantitativa del acuerdo entre jueces, complementado con intervalos de confianza y un análisis cualitativo de las sugerencias (Rodríguez et al., 2016).

Los resultados de la primera ronda de evaluación mostraron un nivel de aceptación generalmente bueno, con promedios de V de Aiken por ítem que oscilaron entre 0.770 y 0.938. Sin embargo, la aplicación de los criterios de Aiken (1985) y Penfield y Giacobbi (2004) permitió identificar ítems con validez baja o aceptable que requerían revisión, como el ítem 3 ($V=0.56$ en claridad) y el ítem 1 ($V=0.67$ en pertinencia). Este diagnóstico preciso fue crucial, ya que orientó las modificaciones específicas. Por ejemplo, la ambigüedad en la redacción del ítem 3, señalada por los expertos, se corrigió eliminando referencias externas confusas y reformulando la alternativa clave para evaluar de manera más directa el proceso de indagación.

El proceso iterativo de revisión y la segunda ronda de evaluación demostraron la efectividad de esta metodología. Los valores finales de la V de Aiken experimentaron una mejora generalizada y consistente. Destaca el caso del ítem 3, cuyo coeficiente de claridad pasó de 0.56 a 0.78, superando el umbral de aceptabilidad. Asimismo, el ítem 1 mejoró su pertinencia de 0.67 a 0.89 tras una reunión de clarificación conceptual con los jueces. Al final del proceso, todos los ítems alcanzaron o superaron el valor de 0.80 en sus promedios finales, ubicándose en las categorías de validez "buena a muy buena" y "excelente". El ítem 2, por ejemplo, alcanzó un promedio de 0.931, lo que indica un acuerdo casi unánime entre los expertos sobre su calidad.

La importancia de reportar intervalos de confianza para la V de Aiken, como recomienda Domínguez-Lara (2015), radica en que proporcionan información sobre la precisión y estabilidad de la estimación. En este estudio, el hecho de que los límites inferiores de estos intervalos se

mantuvieran por encima de 0.70 para la mayoría de los ítems en la evaluación final refuerza la solidez de la evidencia de validez de contenido. Este riguroso proceso metodológico asegura que el instrumento no solo es teóricamente sólido, sino que también sus ítems son claros, pertinentes, relevantes y suficientes para la población objetivo, cumpliendo así con los estándares exigidos para pruebas educativas y psicológicas (AERA et al., 2014; Boateng et al., 2018).

5.3 Evaluación de los resultados piloto y su relación con las problemáticas planteadas

La aplicación piloto del instrumento, si bien reveló problemas psicométricos significativos, constituye una fase esencial de aprendizaje y ajuste, y sus resultados deben analizarse a la luz de las problemáticas planteadas inicialmente.

En primer lugar, los resultados confirman la complejidad de operacionalizar el constructo de curiosidad epistémica, especialmente la dimensión de Privación (Tipo-D), en niños de primaria. Como se planteaba en el marco teórico, la curiosidad es un constructo multifacético y su manifestación está mediada por el desarrollo cognitivo y lingüístico (Berlyne, 1954; Gopnik et al., 2015; Jirout & Klahr, 2012). Los datos piloto son elocuentes: los ítems correspondientes a las subdimensiones B3 ("Estado de ansiedad o insatisfacción") y B4 ("Busca reducir estados indeseables de incertidumbre") mostraron índices de dificultad extremos ($p < .20$), discriminación nula o negativa, y confiabilidad interna inaceptable en los tres niveles evaluados (2°, 3° y 4° básico). Esto sugiere que los conceptos de "ansiedad epistémica" o la "reducción de incógnitas específicas" pueden ser demasiado abstractos o difíciles de autorreportar para niños entre 7 y 10 años, quienes pueden no tener conciencia metacognitiva de estos estados o carecer del vocabulario para expresarlos. Este hallazgo es coherente con la literatura que advierte sobre los desafíos de medir constructos motivacionales complejos en la infancia (McNary, 2023).

En segundo lugar, los resultados resaltan la necesidad crítica de adecuar el lenguaje y el formato del instrumento al desarrollo infantil, una preocupación ya abordada en las entrevistas cognitivas pero que la aplicación a mayor escala puso en evidencia. La baja varianza, la concentración de respuestas en distractores específicos y los coeficientes de confiabilidad insuficientes (α entre .351 y .644) indican que, más allá de la validez de contenido, existen problemas de validez referida al proceso de respuesta (Padilla & Benítez, 2014). Es probable que la redacción, la extensión de las frases o el formato de selección múltiple con cuatro alternativas hayan resultando desafiantes, especialmente para los estudiantes de segundo básico. Esto valida la necesidad, señalada en los antecedentes, de crear instrumentos que no sean meras adaptaciones lingüísticas de escalas para adultos, sino que sean diseñados desde cero considerando las capacidades de la población objetivo (Ahmad & Siew, 2021).

Finalmente, los resultados ofrecen una dirección clara para la revisión del instrumento y aportan evidencia empírica a los vacíos identificados. Si bien el instrumento en su forma actual no es psicométricamente viable, el patrón diferenciado en el funcionamiento de los ítems es informativo. Las dimensiones de Curiosidad de Interés (Tipo-I), particularmente los ítems relacionados con el disfrute de descubrimientos (A3) y la activación ante lo nuevo (A2), mostraron un comportamiento más promisorio, con dificultades adecuadas y, en algunos casos (como el ítem 1 en 4° básico), una discriminación excelente ($r = .694$). Esto sugiere que los aspectos positivos y de "placer por aprender" de la curiosidad podrían ser más accesibles y medibles en esta etapa del desarrollo. Por lo tanto, los resultados no invalidan el modelo teórico de Litman (2008), sino que precisan su aplicabilidad: para el ciclo básico inicial, podría ser necesario priorizar la evaluación de la Curiosidad de Interés o reformular de manera radical los ítems de Privación para hacerlos más concretos y observables.

En conclusión, el proceso de desarrollo y validación, que incluyó la etapa piloto, ha sido exitoso en su propósito de diagnóstico y refinamiento. Los hallazgos confirman la pertinencia de la problemática planteada (la carencia de instrumentos adecuados) y proporcionan un camino empíricamente fundado para superarla. Las próximas etapas deberán centrarse en una reestructuración profunda del instrumento, considerando: 1) la eliminación o reformulación completa de los ítems de Privación (B3 y B4); 2) la simplificación lingüística y el rediseño de distractores en todos los ítems; 3) la posible reducción del número de alternativas de respuesta; y 4) la realización de una nueva prueba piloto, complementada con entrevistas cognitivas, para verificar los ajustes. De este modo, este estudio sienta las bases metodológicas y teóricas robustas para la creación final de una herramienta válida y confiable que contribuya a la investigación y la promoción de la curiosidad epistémica en las aulas de matemática chilena.

VI. CONCLUSIONES, LIMITACIONES Y PROYECCIONES DEL ESTUDIO

6.1 Conclusiones de objetivos y preguntas de investigación

La presente investigación se propuso desarrollar y validar una escala cuantitativa para medir la curiosidad epistémica en matemáticas, específicamente en contextos de razonamiento inferencial informal, en estudiantes de segundo a cuarto básico del sistema escolar chileno. Este proceso se llevó a cabo a través de un diseño metodológico riguroso que incluyó una investigación sistemática, validación de contenido mediante juicio de expertos, entrevistas cognitivas y una aplicación piloto, con el fin de generar un instrumento válido, confiable y contextualizado.

6.1.1 Conclusiones del objetivo general

El objetivo general del estudio fue “Diseñar y validar una escala cuantitativa, válida y confiable, para medir la curiosidad epistémica de tipo I/D en contextos de razonamiento inferencial informal en matemática, en estudiantes que cursan desde segundo a cuarto básico en Chile”

Tras el proceso de desarrollo y validación, se concluye que el objetivo general se cumplió parcialmente. Se logró diseñar un instrumento de ocho ítems, estructuralmente basado en el modelo bidimensional de Litman (2008), y contextualizado en escenarios de razonamiento inferencial informal alineados con el currículum nacional. La validación de contenido, ejecutada mediante el juicio de seis expertos y entrevistas cognitivas con diez estudiantes, demostró que la escala alcanza niveles adecuados de claridad, pertinencia, relevancia y suficiencia tras una revisión iterativa, con valores de la V de Aiken entre 0.80 y 1.00 para la mayoría de los criterios (Rodríguez, Gil y García, 2016; Aracena et al., 2024). Sin embargo, los resultados de la aplicación piloto, que evaluó el funcionamiento psicométrico en muestras de segundo (n=21), tercero (n=24) y cuarto

(n=30) básico, revelaron que el instrumento en su versión actual no alcanza los estándares mínimos de confiabilidad y validez de constructo necesarios para su uso generalizado. La consistencia interna según el α de Cronbach fue “inaceptable” en todos los niveles puntuaciones entre .351 y .644, y se identificaron problemas críticos en la discriminación y dificultad de los ítems, especialmente aquellos correspondientes a la dimensión de Curiosidad por Privación (Tipo-D). Por lo tanto, si bien se completó el diseño y una validación de contenido robusta, la validación psicométrica evidenció la necesidad de una revisión profunda del instrumento antes de que pueda considerarse una herramienta confiable para la medición del constructo en esta población.

6.1.2 Conclusiones del objetivo específico uno

El primer objetivo específico planteado fue “Identificar instrumentos existentes para medir la curiosidad epistémica en contextos matemáticos”.

Para dar respuesta a este objetivo, se realizó una revisión exploratoria sistemática en abril de 2025, consultando bases de datos como Web of Science, Scielo, Scopus y Google Académico. Los resultados de esta revisión fueron concluyentes: existe un vacío crítico en la literatura respecto a instrumentos específicos para medir la curiosidad epistémica en el dominio matemático durante la infancia (Ahmad y Siew, 2021; Jirout, 2020). Si bien se identificaron instrumentos genéricos, como la Escala de Curiosidad Epistémica de Litman validada al español por Sánchez et al. (2023), estos fueron diseñados para población adulta y presentan limitaciones metodológicas (como desequilibrio de género y falta de muestreo probabilístico) que los hacen inaplicables para estudiantes de primaria (Bengart et al., 2018). La revisión confirmó la necesidad y pertinencia del presente estudio, al no encontrar escalas que captarán las manifestaciones situadas de la curiosidad

en escenarios de razonamiento inferencial informal propios del aprendizaje matemático inicial en el contexto chileno (Ministerio de Educación de Chile, 2018).

6.1.3 Conclusiones del objetivo específico dos

El segundo objetivo específico fue “Construir una escala cuantitativa para medir la curiosidad epistémica, en matemática, en estudiantes que cursan segundo a cuarto básico en Chile”

Este objetivo se cumplió de manera satisfactoria. Se construyó una escala preliminar de ocho ítems, operacionalizando el modelo teórico de Litman (2008) en sus dos dimensiones principales: Curiosidad de Interés (Tipo-I) y Curiosidad por Privación (Tipo-D). Cada dimensión se desglosó en cuatro características, asignando un ítem por cada una de ellas (ver Tabla 1). La construcción fue un proceso colaborativo que integró a investigadores en formación y un experto en didáctica de la matemática, utilizando como guía una propuesta de enseñanza para el razonamiento inferencial informal (Estrella, 2025). Un aporte distintivo fue la contextualización de cada ítem en escenarios auténticos y ecológicamente válidos, alineados explícitamente con los Objetivos de Aprendizaje del currículum nacional chileno desde 2° a 4° básico (ver Tabla 2). Esto aseguró que el instrumento midiera el constructo no como un rasgo general, sino tal como se manifiesta en situaciones específicas de exploración matemática relevantes para la experiencia escolar de los estudiantes, abordando así la carencia identificada en la revisión de literatura.

6.1.4 Conclusiones del objetivo específico tres

El tercer objetivo específico fue: Validar la escala cuantitativa para medir la curiosidad epistémica, en matemática, en estudiantes que se encuentran cursando segundo, tercero y cuarto básico en Chile.

La validación del instrumento se abordó en dos etapas principales: validación de contenido y una evaluación piloto de sus propiedades psicométricas. Respecto a la validez de contenido, el proceso fue exitoso. El juicio de 6 expertos en dos rondas iterativas permitió refinar significativamente los ítems. Tras las modificaciones, siete de los ocho ítems alcanzaron valores de V de Aiken superiores a 0.80 en los criterios de claridad, pertinencia y relevancia, y la suficiencia de la escala en su conjunto fue validada (Rodríguez et al., 2016). Complementariamente, las entrevistas cognitivas a 10 estudiantes proporcionaron evidencia crucial sobre la validez referida al proceso de respuesta (Padilla y Benítez, 2014; Willis, 2015). Los estudiantes comprendieron la tarea, se mostraron comprometidos (engagement) y las entrevistas permitieron identificar y corregir problemas lingüísticos y conceptuales específicos (por ejemplo, en los ítems 3 y 8), fortaleciendo la claridad y pertinencia percibida.

No obstante, los resultados de la aplicación piloto realizada a 75 estudiantes indican que la validación psicométrica del instrumento en su forma actual es insuficiente. Si bien se observó una mejora progresiva en los indicadores desde segundo a cuarto básico, en ningún nivel se alcanzaron los estándares psicométricos mínimos. La confiabilidad medida con alfa de Cronbach fue consistentemente baja (segundo: $\alpha=.502$; tercero: $\alpha=.351$; cuarto: $\alpha=.644$) (Cortina, 1993; Cronbach, 1951). El análisis ítem a ítem reveló un patrón claro: los reactivos de la dimensión Curiosidad por Privación (Tipo-D), especialmente aquellos relacionados con la "ansiedad por incertidumbre" (B3) y la "reducción de incógnitas" (B4), presentaron índices de dificultad extremos ($p < .20$), discriminación nula o negativa, y distractores que capturaron la mayoría de las respuestas (Haladyna, 2004; Nunnally y Bernstein, 1994). Esto sugiere que estos constructos más abstractos y de valencia afectiva negativa podrían no ser apropiados para ser medidos mediante auto-reporte en niños de esta edad, o que su operacionalización en los ítems actuales no es

adecuada. En contraste, los ítems de la dimensión Curiosidad de Interés (Tipo-I) mostraron un funcionamiento más promisorio, con dificultades dentro del rango aceptable en su mayoría.

Por lo tanto, se concluye que el proceso de validación de contenido fue robusto y exitoso, generando un instrumento con una sólida base teórica y contextual. Sin embargo, la validación psicométrica a través de la aplicación piloto evidenció limitaciones estructurales significativas que impiden, por el momento, considerar la escala como un instrumento confiable y válido para su aplicación diagnóstica o investigativa a gran escala. Esta fase piloto cumplió su propósito esencial: detectar fallas críticas y proporcionar un diagnóstico detallado para una reestructuración informada del instrumento (Boateng et al., 2018).

6.2 Limitaciones y proyecciones del estudio

6.2.1 Limitaciones del estudio

El desarrollo y validación de este instrumento se enfrentó a varias limitaciones que es importante reconocer para una correcta interpretación de los hallazgos y para orientar futuras investigaciones:

1. Limitaciones muestrales y de acceso: El proceso de reclutamiento de la muestra enfrentó dificultades logísticas significativas. Se contactó de manera formal a cuatro establecimientos educativos distintos antes de conseguir una participación favorable, siendo la falta de respuesta de estos una barrera inesperada que limitó las opciones disponibles. Como consecuencia, la muestra para la aplicación piloto fue restringida geográficamente a un solo establecimiento en la comuna de Los Ángeles, Chile. Si bien este tipo de muestreo es común en las fases iniciales de validación instrumental, esta

limitación geográfica y contextual restringe la diversidad de la muestra y, por ende, la posibilidad de generalizar los resultados a otros contextos regionales, culturales o socioeconómicos dentro de la región del Biobío adicionalmente, el tamaño de la muestra piloto, aunque técnicamente adecuado para una primera evaluación psicométrica, fue modesto (75 estudiantes distribuidos en tres cursos). Un tamaño muestral mayor habría proporcionado estimaciones estadísticas más estables y precisas, particularmente para los análisis de confiabilidad y los índices de discriminación por ítem (Boateng et al., 2018; Field, 2018)

2. Limitaciones metodológicas en la operacionalización del constructo: La principal limitación identificada radica en la dificultad para operacionalizar la dimensión de Curiosidad por Privación (Tipo-D) para la población objetivo. Los ítems correspondientes a las subdimensiones B3 ("Estado de ansiedad o insatisfacción") y B4 ("Busca reducir estados indeseables de incertidumbre") mostraron un desempeño psicométrico deficiente. Esto podría deberse a que estos estados cognitivo-afectivos, que involucran la percepción de una "brecha de información" y el deseo de reducir un estado aversivo (Loewenstein, 1994; Litman, 2008), son demasiado abstractos o complejos para que niños de 7 a 10 años los autorreporten de manera fiable mediante un cuestionario de lápiz y papel. La teoría del desarrollo cognitivo sugiere que, aunque en esta edad se consolida el pensamiento lógico concreto, la introspección y verbalización de estados emocionales complejos ligados al conocimiento pueden ser limitadas (Inhelder y Piaget, 1958; Gopnik et al., 2015).
3. Limitaciones en el formato de respuesta: El instrumento utilizó un formato de respuesta de opción múltiple con cuatro alternativas basadas en situaciones. Los resultados piloto,

especialmente en segundo básico, mostraron distribuciones de respuestas concentradas en pocas alternativas y baja varianza, lo que podría indicar que el formato no es el más adecuado para las edades más tempranas. Para niños pequeños, formatos más visuales (escalas de caras y/o elección entre pocas opciones claramente ilustradas) podrían ser más apropiados y reducir el ruido asociado a la comprensión lectora (Boateng et al., 2018).

4. Alcance de la validación: El estudio avanzó hasta la fase de aplicación piloto y análisis de las propiedades psicométricas preliminares. No se llevó a cabo un análisis factorial confirmatorio (AFC) con una muestra más grande, el cual es necesario para verificar empíricamente la estructura bidimensional postulada por la teoría (Ferrando et al., 2022; Aracena et al., 2024). Otra limitación importante de este estudio es que no se evaluó la validez convergente y divergente del instrumento, es decir, su relación con otras medidas teóricamente vinculadas o no vinculadas al constructo, como por ejemplo el rendimiento académico en matemáticas o subescalas de curiosidad, interés y disfrute de esta área. Por ello, la evidencia de validez de constructo se limita principalmente a la validez de contenido y a los indicadores internos del instrumento. Esto se debió principalmente al tiempo acotado disponible para el desarrollo del estudio, lo que llevó a priorizar las fases fundamentales de diseño teórico, validación de contenido y evaluación psicométrica preliminar.

6.2.2 Proyecciones del estudio

A partir de los hallazgos y limitaciones identificadas, se delinean las siguientes proyecciones para continuar el desarrollo de esta línea de investigación:

Revisión profunda y reestructuración del instrumento: El paso inmediato más crítico es una revisión integral del instrumento fundamentada en los resultados psicométricos de la fase piloto. Esto implica, eliminar o reformular radicalmente los ítems de la dimensión Tipo-D (Privación), especialmente los correspondientes a B3 y B4. Podría explorarse una reformulación que utilice un lenguaje más concreto, referido a situaciones de "no entender" o "querer saber sí o sí" que sean más identificables para los niños. También se sugiere reformular los ítems de la dimensión Tipo-I (Interés) que mostraron baja discriminación o dificultad extrema, simplificando el lenguaje, acortando las frases y asegurando que las alternativas sean claramente diferenciables. Por otro lado se proyecta rediseñar los distractores de todos los ítems, sometiéndolos a un nuevo juicio de expertos que incluya docentes de primer ciclo, con el fin de asegurar que las alternativas resulten plausibles sin llegar a ser más atractivas que la opción correcta. Y por último, identificar la posibilidad de cambiar el formato de respuesta hacia opciones más visuales o binarias para los cursos inferiores (2° y 3° básico), manteniendo un formato más textual solo para 4° básico.

Otra proyección es un nuevo ciclo de validación con una muestra ampliada y representativa: Una vez reformulado, el instrumento debe someterse a un nuevo ciclo de validación que incluya: Una nueva aplicación piloto con una muestra más grande y estratificada por curso, que permita realizar análisis factoriales exploratorios y confirmatorios robustos (AFC) para verificar la estructura dimensional del instrumento revisado. La evaluación de la validez de criterio, correlacionando las puntuaciones en la escala con medidas de rendimiento en matemáticas o con observaciones docentes sobre comportamientos curiosos en el aula, también la estimación de la confiabilidad test-retest para evaluar la estabilidad temporal de las mediciones.

Y por último proyecciones para la investigación y la práctica educativa tales como, la investigación aplicada: Una vez validado, el instrumento podría emplearse en estudios que exploren la relación entre la curiosidad epistémica matemática y variables como el logro académico, la autorregulación del aprendizaje, la autoeficacia o la efectividad de metodologías didácticas específicas (por ejemplo, el Estudio de Clases) en el contexto chileno (Jirout, 2020; von Stumm et al., 2011). También tomar el instrumento como una herramienta diagnóstica para docentes: El instrumento final, debidamente validado, podría convertirse en una herramienta diagnóstica valiosa para docentes de primer ciclo, permitiéndoles identificar perfiles de curiosidad en sus estudiantes (predominio de Tipo-I o Tipo-D) y, en consecuencia, diseñar intervenciones pedagógicas diferenciadas que fomenten la exploración, la formulación de preguntas y la persistencia ante desafíos matemáticos (MINEDUC, 2018).

En definitiva, este estudio constituye un primer paso sólido y rigurosamente fundamentado hacia la medición de la curiosidad epistémica en matemáticas en la educación básica chilena. Aunque la versión actual del instrumento requiere importantes ajustes, el proceso ha generado conocimiento valioso sobre las dificultades de operacionalizar este constructo complejo en población infantil y ha establecido una metodología clara para su desarrollo. Las proyecciones delineadas apuntan a transformar este prototipo en una herramienta psicométricamente robusta con un gran potencial para enriquecer tanto la investigación educativa como la práctica pedagógica en Chile.

REFERENCIAS

- Agencia de Calidad de la Educación. (2023). *Resultados Educativos 2022*. Ministerio de Educación de Chile. <https://app.powerbi.com/view?spm=2b75ac3d.2ef5001f.0.0.2377c921BKHvYP&r=eyJrIjoiYzQyOGU1ZjgtZDk1Ni00MjNhLTg3MjYtZGQ5MDE0MDE0ZTNkIiwidCI6IjgwZGU4Zjg3LTFEYmZgtNGQ1NS05M2I2LWMyZDFmMmYzOWY1ZiIsImMiOjR9&pageName=98668c76520606a4ad22>
- Ahmad, J., y Siew, N. M. (2021). Curiosity towards STEM education: A questionnaire for primary school students. *Journal of Baltic Science Education*, 20(3), 455-467. [10.33225/jbse/21.20.289](https://doi.org/10.33225/jbse/21.20.289)
- American Educational Research Association (AERA), American Psychological Association (APA), y National Council on Measurement in Education (NCME). (2014). *Standards for educational and psychological testing*. AERA. <https://www.testingstandards.net/>
- Aracena, M., Gutiérrez, J., López, M., Valenzuela, C., Cuadra, M., Oyanadel, V., Castro, P., y González, P. (2024). Adaptación y validación de la Student Teacher Professional Identity Scale en estudiantes de pedagogía chilenos. *Interdisciplinaria. Revista de Psicología y Ciencias Afines*, 41(1), 1-19. <https://doi.org/10.16888/interd.2024.41.1.9>
- Bahamondes, P., Lecaros, J., Leppe, J., Mattar, G., y Merino, C. (2017). Encuesta sobre implementación de investigación en centros educacionales en Chile: Cumplimientos de aspectos regulatorios, éticos y legales. *Acta Bioethica*, 23(2), 365-374. <https://doi.org/10.4067/S1726-569X2017000200365>

- Bandalos, D. L. (2021). *Measurement theory and applications for the social sciences*. Guilford Publications.
- Baumeister, R. F. (2005). *The cultural animal: Human nature, meaning, and social life*. Oxford University Press.
<https://global.oup.com/academic/product/the-cultural-animal-9780195167030>
- Bengart, P., Sarstedt, M., Shaltoni, A. M., y Lehmann, S. (2018). The use of sampling methods in advertising research: A gap between theory and practice. *International Journal of Advertising*, 37(4), 650-663. <https://doi.org/10.1080/02650487.2017.1348329>
- Berg, C. A., y Sternberg, R. J. (1985). Response to novelty: Continuity versus discontinuity in the developmental course of intelligence. *Advances in Child Development and Behavior*, 19, 1–47. [https://doi.org/10.1016/s0065-2407\(08\)60387-0](https://doi.org/10.1016/s0065-2407(08)60387-0)
- Berlyne, D. E. (1954). A theory of human curiosity. *British Journal of Psychology*, 45(3), 180–191. <https://doi.org/10.1111/j.2044-8295.1954.tb01243.x>
- Berlyne, D. E. (1966). Curiosity and exploration. *Science*, 153(3731), 25–33. <https://doi.org/10.1126/science.153.3731.25>
- Berlyne, D. E. (1971). What next? Concluding summary. In H. I. Day, D. E. Berlyne y D. E. Hunt (Eds.), *Intrinsic motivation: A new direction in education* (pp. 186–196). Holt, Rinehart and Winston of Canada. <https://www.worldcat.org/title/Intrinsic-motivation-a-new-direction-in-education/oclc/832933>

Biblioteca del Congreso Nacional de Chile. (2006, 22 de septiembre). *Ley 20120: Sobre la investigación científica en el ser humano, su genoma, y prohíbe la clonación humana.*

[Ley Chile - Ley 20120 - Biblioteca del Congreso Nacional](#)

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., y Young, S. L. (2018).

Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health*, 6, 149.

<https://doi.org/10.3389/fpubh.2018.00149>

Bonwell, C. C., y Eison, J. A. (1991). *Active learning: Creating excitement in the classroom* (ASHE-ERIC Higher Education Report No. 1). The George Washington University, School of Education and Human Development.

<https://eric.ed.gov/?id=ED336049>

Bosch, N., y D'Mello, S. K. (2017). Affective computing for education: A review. *IEEE Transactions on Affective Computing*, 9(4), 405-423.

<https://doi.org/10.1109/TAFFC.2017.2744352>

Chi, M. T. H., De Leeuw, N., Chiu, M.-H., y LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, 18(3), 439-477.

https://doi.org/10.1207/s15516709cog1803_3

Chi, M. T. H., y Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, 49(4), 219-243.

<https://doi.org/10.1080/00461520.2014.965823>

- Chouinard, M. M., Harris, P. L., y Maratsos, M. P. (2020). Children's questions: A mechanism for cognitive development. *Monographs of the Society for Research in Child Development*, 72(1), Serial No. 286. <https://doi.org/10.1111/j.1540-5834.2007.00412.x>
- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, 78(1), 98–104. <https://doi.org/10.1037/0021-9010.78.1.98>
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297-334. <https://doi.org/10.1007/BF02310555>
- DiStefano, C., y Finney, S. J. (2013). Nonnormal and categorical data in structural equation modeling. In G. R. Hancock y R. O. Mueller (Eds.), *Structural equation modeling: A second course* (2nd ed., pp. 439–492). Information Age Publishing.
- Domínguez-Lara, S. (2015). ¿Por qué es importante reportar intervalos de confianza al usar la V de Aiken? *Acción Psicológica*, 16(2), 1326-1328. <https://doi.org/10.5944/ap.16.2.25819>
- Ellis, A. B. (2007). Connections between generalizing and justifying: Students' reasoning with linear relationships. *Journal for Research in Mathematics Education*, 38(3), 194–229.
DOI: [10.2307/30034866](https://doi.org/10.2307/30034866)
- Engel, S. (2015). *The hungry mind: The origins of curiosity in childhood*. Harvard University Press. <https://www.hup.harvard.edu/catalog.php?isbn=9780674736757>
- Estrella, S. (2025). What does preparing our boys and girls to interact with artificial agents mean? Perspectives from Early Statistical Education in Today's Classroom. In C. Cornejo, P. Felmer, D. M. Gómez, P. Dartnell, P. Araya, A. Peri, y V. Randolph (Eds.), *Proceedings*

of the 48th Conference of the International Group for the Psychology of Mathematics Education: General Contributions (pp. 55-69). International Group for the Psychology of Mathematics Education.

Ferrando, P. J., Lorenzo-Seva, U., Hernández-Dorado, A., y Muñiz, J. (2022). Decálogo para el análisis factorial de los ítems de un test. *Psicothema*, 34(1), 7-17.
<https://doi.org/10.7334/psicothema2021.456>

Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). Sage Publications.
<https://uk.sagepub.com/en-gb/eur/discovering-statistics-using-ibm-spss-statistics/book259067>

Giedd, J. N., Raznahan, A., Alexander-Bloch, A., Schmitt, E., Gogtay, N., y Rapoport, J. L. (2015). Child psychiatry branch of the National Institute of Mental Health longitudinal structural magnetic resonance imaging study of human brain development. *Neuropsychopharmacology*, 40(1), 43-49.
<https://doi.org/10.1038/npp.2014.236>

George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference. 11.0 update* (4th ed.). Allyn & Bacon.

George, D., & Mallery, P. (2019). *IBM SPSS Statistics 26 Step by Step: A Simple Guide and Reference* (16th ed.). Routledge. <https://doi.org/10.4324/9780429056765>

Gopnik, A., Meltzoff, A. N., y Kuhl, P. K. (2015). *The scientist in the crib: Minds, brains, and how children learn.* William Morrow Paperbacks.

<https://www.harpercollins.com/products/the-scientist-in-the-crib-alison-gopnik-andrew-n-meltzoff-patricia-k-kuhl?variant=32122717306914>

Haladyna, T. M. (2004). *Developing and validating multiple-choice test items* (3rd ed.). Lawrence Erlbaum Associates.

Hartung, F.-M., Thieme, P., Wild-Wall, N., y Hell, B. (2024). Being nosy and smart: The relationship between curiosity, fluid intelligence, and knowledge. *Cognition and Emotion*, 38(4), 567-581. <https://doi.org/10.1027/1614-0001/a000372>

Haynes, S. N., Richard, D. C. S., y Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment*, 7(3), 238–247. <https://doi.org/10.1037/1040-3590.7.3.238>

Heaviside, S. (1993). *Perspectivas de los docentes de kínder en escuelas públicas sobre la preparación de los niños para la escuela: informe de un contratista*. Departamento de Educación de los Estados Unidos, Oficina de Investigación y Mejora Educativa. ISBN: 016042030X

Heaviside, S., Farris, E., y Carroll, D. C. (2015). *Public school kindergarten teachers' views on children's readiness for school* (NCES Report No. 2015-097). National Center for Education Statistics. <https://nces.ed.gov/pubs2015/2015097.pdf>

Hidi, S., y Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist*, 41(2), 111–127. https://doi.org/10.1207/s15326985ep4102_4

- Inhelder, B., y Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence*. Basic Books. (*Libro clásico. Enlace a WorldCat*): <https://www.worldcat.org/title/growth-of-logical-thinking-from-childhood-to-adolescence/oclc/234708>
- Jirout, J. J. (2020). Supporting early scientific thinking through curiosity. *Frontiers in Psychology*, *11*, 1717. <https://doi.org/10.3389/fpsyg.2020.01717>
- Jirout, J. J., y Klahr, D. (2012). Children's scientific curiosity: In search of an operational definition of an elusive concept. *Developmental Review*, *32*(2), 125–160. <https://doi.org/10.1016/j.dr.2012.04.002>
- Jirout, J. J., y Zimmerman, C. (2015). Development of science process skills in the early childhood years. In K. C. Trundle y M. Saçkes (Eds.), *Research in early childhood science education* (pp. 143–165). Springer. https://doi.org/10.1007/978-94-017-9505-0_7
- Jirout, J. J., Zumbunn, S., Evans, N. S., y Vitiello, V. E. (2022). Curiosity in classrooms: An examination of curiosity promotion and suppression in preschool math and science classrooms. *Frontiers in Psychology*, *13*, 875161. <https://doi.org/10.3389/fpsyg.2022.875161>
- Kashdan, T. B., Disabato, D. J., Goodman, F. R., y McKnight, P. E. (2020). The five-dimensional curiosity scale: Capturing the bandwidth of curiosity and identifying four unique subgroups of curious people. *Journal of Research in Personality*, *83*, 103955. <https://doi.org/10.1016/j.jrp.2017.11.011>
- Kidd, C., y Hayden, B. Y. (2015). The psychology and neuroscience of curiosity. *Neuron*, *88*(3), 449–460. <https://doi.org/10.1016/j.neuron.2015.09.010>

- King, A. (1994). Guiding knowledge construction in the classroom: Effects of teaching children how to question and how to explain. *American Educational Research Journal*, 31(2), 338-368. <https://doi.org/10.3102/00028312031002338>
- Klahr, D., Matlen, B., y Jirout, J. (2013). Children as scientific thinkers. In G. J. Feist y M. E. Gorman (Eds.), *Handbook of the psychology of science* (pp. 223–248). Springer. file:///C:/Users/56965/Downloads/klahr.matlen.jirout1.pdf
- Legare, C. H. (2014). The contributions of explanation and exploration to children's scientific reasoning. *Child Development Perspectives*, 8(2), 101–106. <https://doi.org/10.1111/cdep.12070>
- Litman, J. A. (2005). Curiosity and the pleasures of learning: Wanting and liking new information. *Cognition and Emotion*, 19(6), 793–814. <https://doi.org/10.1080/02699930541000101>
- Litman, J. A. (2007). Curiosity as a feeling of interest and feeling of deprivation: The I/D model of curiosity. In P. R. Zeilek (Ed.), *Issues in the psychology of motivation* (pp. 149–156). Nova Science Publishers. (Libro de editorial pequeña. Enlace a WorldCat): <https://www.worldcat.org/title/Issues-in-the-psychology-of-motivation/oclc/123539663>
- Litman, J. A. (2008). Interest and deprivation factors of epistemic curiosity. *Personality and Individual Differences*, 44(7), 1585–1595. <https://doi.org/10.1016/j.paid.2008.01.014>
- Litman, J. A., y Jimerson, T. L. (2004). The measurement of curiosity as a feeling of deprivation. *Journal of Personality Assessment*, 82(2), 147–157. https://doi.org/10.1207/s15327752jpa8202_3

- Litman, J. A., y Spielberger, C. D. (2003). Measuring epistemic curiosity and its diversive and specific components. *Journal of Personality Assessment*, 80(1), 75–86. https://doi.org/10.1207/S15327752JPA8001_16
- Littman, J. A., Piotrowski, J. T., y Valkenburg, P. (2014). Measuring epistemic curiosity in young children. *Personality and Individual Differences*, 60(Supplement), S8. <https://doi.org/10.1016/j.paid.2013.07.173>
- Liu, J., Tahri, D., y Qiang, F. (2024). How does active learning pedagogy shape student curiosity? A multisite mediation study of student engagement among 45,972 children. *Journal of Intelligence*, 12(6), 59. <https://doi.org/10.3390/jintelligence12060059>
- Loewenstein, G. (1994). The psychology of curiosity: A review and reinterpretation. *Psychological Bulletin*, 116(1), 75–98. <https://doi.org/10.1037/0033-2909.116.1.75>
- Loyens, S. M. M., y Gijbels, D. (2008). Understanding the effects of constructivist learning environments: Introducing a multi-directional approach. *Instructional Science*, 36(5-6), 351-357. <https://doi.org/10.1007/s11251-008-9059-4>
- Martin, T., y Schwartz, D. L. (2004). Inventing to prepare for future learning: The hidden efficiency of encouraging original student production in statistics instruction. *Cognition and Instruction*, 22(2), 129–184. https://doi.org/10.1207/s1532690xci2202_1
- McNary, L. (2023). Curiosity reanalysis: A conceptual review to improve measurement. *Educational Psychology Review*, 35(2), 55. <https://doi.org/10.1007/s10648-023-09777-7>

- Menning, C. (2019). Curiosity in the classroom: Promoting student engagement through teacher-student interaction. *Teaching and Teacher Education*, 77, 297-305. <https://doi.org/10.1016/j.tate.2018.10.012>
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741-749. <https://doi.org/10.1037/0003-066X.50.9.741>
- Ministerio de Educación de Chile. (2018). *Bases Curriculares 1° a 6° Básico*. Unidad de Currículum y Evaluación. https://www.curriculumnacional.cl/sites/default/files/newtenberg/614/articles-22394_bases.pdf
- Ministerio de Salud de Chile. (2006). *Reglamento de la Ley N° 20.120 sobre la Investigación Científica en el Ser Humano, su Genoma, y Prohíbe la Clonación Humana*. Decreto Supremo N° 114. <https://bcn.cl/2w1j1> (Enlace corto oficial al reglamento)
- Mussel, P. (2022). Curiosity and cognitive performance: A meta-analysis. *Personality and Social Psychology Review*, 26(4), 361-383. <https://doi.org/10.3390/jintelligence10020023>
- Nunnally, J. C., y Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.
- OECD. (2023). *PISA 2022 Results (Volume I): The State of Learning and Equity in Education*. OECD Publishing. <https://doi.org/10.1787/53f23881-en>

- Oudeyer, P.-Y., Gottlieb, J., y Lopes, M. (2016). Intrinsic motivation, curiosity, and learning: Theory and applications in educational technologies. *Progress in Brain Research*, 229, 257–284. <https://doi.org/10.1016/bs.pbr.2016.05.005>
- Padilla, J. L., y Benítez, I. (2014). Validity evidence based on response processes. *Psicothema*, 26(1), 136-144. <https://doi.org/10.7334/psicothema2013.259>
- Pekrun, R., Frenzel, A. C., Goetz, T., y Perry, R. P. (2017). The control-value theory of achievement emotions: An integrative approach to emotions in education. In R. Pekrun y L. Linnenbrink-Garcia (Eds.), *Handbook of emotions in education* (pp. 120-141). Routledge.
<https://www.routledge.com/Handbook-of-Emotions-in-Education/Pekrun-Linnenbrink-Garcia/p/book/9780367209920>
- Pekrun, R., Vogl, E., Muis, K. R., y Sinatra, G. M. (2018). Measuring emotions during epistemic activities: The Epistemically Related Emotion Scale. *Contemporary Educational Psychology*, 52, 160-169. <https://doi.org/10.1016/j.cedpsych.2018.01.009>
- Renner, B. (2006). Curiosity about people: The development of a social curiosity measure in adults. *Journal of Personality Assessment*, 87(3), 305–316.
https://doi.org/10.1207/s15327752jpa8703_11
- Rodríguez, P., Gil, L., y García, E. (2016). Validez de contenido usando la V de Aiken con intervalos de confianza: Aportes metodológicos. *Revista Electrónica de Metodología Aplicada*, 21(1), 79–97.
<https://www.sap.org.ar/docs/publicaciones/archivosarg/2016/v114n3a33.pdf>

- Sánchez, M., Álvarez-Montero, F., y Reyes-Sosa, H. (2023). Evaluación de las propiedades psicométricas de la Escala de Curiosidad Epistémica de Litman en hispanohablantes. *Acta Colombiana de Psicología*, 26(2), 114–127. <https://doi.org/10.14718/ACP.2023.26.2.10>
- Sarac, S., Mede, E., y Sarac, S. (2023). Epistemic curiosity scale for young children: A Turkish adaptation of the scale. *Estudios de Psicología*, 43(1), 41–58. <https://doi.org/10.26650/SP2021-05051>
- Schmalt, H.-D., y Schneider, K. (2009). *Motivation*. Kohlhammer Verlag. (Libro en alemán. Enlace a la editorial): <https://shop.kohlhammer.de/motivation-9783170188413>
- Sellbom, M., y Tellegen, A. (2019). Factor analysis in psychological assessment research: Common pitfalls and recommendations. *Psychological Assessment*, 31(12), 1428-1441. <https://doi.org/10.1037/pas0000623>
- Shah, P., Weeks, H. M., Richards, B., Kaciroti, N., y Lumeng, J. C. (2018). Early childhood curiosity and kindergarten reading and math academic achievement. *Pediatric Research*, 84(3), 380-386. <https://doi.org/10.1038/s41390-018-0039-3>
- Siegler, R. S., y Braithwaite, D. W. (2017). Numerical development. *Annual Review of Psychology*, 68, 187-213. <https://doi.org/10.1146/annurev-psych-010416-044101>
- Spielberger, C. D., y Starr, L. M. (1994). Curiosity and exploratory behavior. In H. F. O'Neil Jr. y M. Drillings (Eds.), *Motivation: Theory and research* (pp. 221–243). Lawrence Erlbaum

Associates. (Libro antiguo. Enlace a WorldCat):

<https://www.worldcat.org/title/Motivation-theory-and-research/oclc/28799692>

Stylianides, G. J., y Stylianides, A. J. (2017). Advancing research in mathematics education: The case of proof. In J. Cai (Ed.), *Compendium for research in mathematics education* (pp. 144–166). National Council of Teachers of Mathematics.

<https://pubs.nctm.org/view/book/9780873537110/9780873537110.xml>

Tabachnick, B. G., y Fidell, L. S. (2019). *Using multivariate statistics* (7th ed.). Pearson.

<https://www.pearson.com/en-us/subject-catalog/p/using-multivariate-statistics/P200000006419/9780137504385>

Tang, S., Xu, T., Jin, L., Ji, L., Chen, Q., y Qiu, J. (2024). *Validation of the I- and D-type epistemic curiosity scale among young Chinese children and implications on early curiosity nurture.*

BMC Psychology, 12, 795. <https://doi.org/10.1186/s40359-024-02286-4>

Vilhunen, J., Malmberg, J., y Järvelä, S. (2022). Epistemic emotions and motivation in collaborative learning. *Learning and Instruction*, 82, 101659.

<https://doi.org/10.1016/j.learninstruc.2022.101659>

Wahl, R. A., Schonbrodt, F. D., y van Bork, R. (2022). How to test a theory, not a measure: A primer on the theoretical status of latent variables. *Advances in Methods and Practices in Psychological Science*, 5(2), 1-19. <https://doi.org/10.1177/25152459221083125>

Wiehler, A., Branzoli, F., Adanyeguh, I., Mochel, F., y Pessiglione, M. (2022). A neuro-metabolic account of why daylong cognitive work alters the control of economic decisions. *Current*

Biology,

32(16),

3564-3575.e5.

<https://doi.org/10.1016/j.cub.2022.07.010>

Willis, G. B. (2015). *Analysis of the cognitive interview in questionnaire design*. Oxford University

Press. <https://psycnet.apa.org/record/2015-02424-000>

ANEXOS

7.1 ANEXO 1: MANUAL DE PROTOCOLO PARA LA APLICACIÓN DE LA ENTREVISTA COGNITIVA

MANUAL DE PROTOCOLO PARA LA APLICACIÓN DE LA ENTREVISTA COGNITIVA: EVALUACIÓN DE LA CURIOSIDAD EPISTÉMICA EN RAZONAMIENTO INFERENCIAL INFORMAL (CE-RII)

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 ayudarás 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 o en las alternativas 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?"
- "¿Cómo lo vas a resolver?"

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?" (señalar alguna opción y preguntar)

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 'curiosidad', 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.

7.2 ANEXO 2: INSTRUMENTO DE VALIDACIÓN

INSTRUMENTO DE VALIDACIÓN: CURIOSIDAD EPISTÉMICA EN RAZONAMIENTO INFERENCIAL INFORMAL (CE-RII)

Estimado/a juez experto/a,

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 Candia y las tesis de la carrera de educación general básica Valeria Albial y Guislene Cifuentes. Desde ya, le agradecemos su participación en la validación de constructo de este instrumento.

El objetivo de este estudio es validar un instrumento para medir la curiosidad epistémica en estudiantes de 7 a 10 años en contextos de razonamiento inferencial informal, considerando dos dimensiones: Curiosidad de Interés (tipo I) y Curiosidad por Privación (tipo D).

A continuación, encontrará los ítems organizados por dimensión y características. Le solicitamos que evalúe cada ítem según los criterios de claridad, pertinencia y relevancia, utilizando una escala Likert de 1 a 4. Al final de cada dimensión, valoraremos la suficiencia de los ítems propuestos.

Tabla 1.

DIMENSIONES	CARACTERÍSTICAS	CANTIDAD DE ÍTEM	CARACTERÍSTICAS GENERALES
A. Curiosidad de Interés (I)	A1. Motivación intrínseca para obtener información	1	Curiosidad de Interés (I) Relacionadas con el deseo de adquirir conocimiento por interés genuino, disfrute y motivación intrínseca (Litman, 2008). Es el deseo de adquirir nuevo conocimiento motivado por la anticipación de la recompensa emocional positiva que genera el proceso de aprendizaje en sí mismo, induciendo sentimientos de interés y disfrute (Litman, 2005; Litman y Jimerson, 2004).
	A2. Se activa al conocer algo nuevo	1	
	A3. Disfruta de nuevos descubrimientos	1	
	A4. Buscar nueva información para estimular sentimientos positivos	1	
B. Curiosidad por Privación (D)	B1. Placer intrínseco y preocupaciones extrínsecas	1	Curiosidad por Privación (D) Relacionadas con la búsqueda de conocimiento para reducir estados de incertidumbre o ansiedad por falta de información (Litman, 2019). Se manifiesta cuando los individuos sienten que carecen de información necesaria; este déficit de conocimiento percibido produce el sentimiento de 'privación' asociado con la curiosidad. (Litman 2005). Enfocada en el desempeño o rendimiento , para las que la corrección, precisión y relevancia del conocimiento recién adquirido son aspectos fundamentales (Litman, 2008).
	B2. Se estimula cuando se carece de información.	1	
	B3. Estado de ansiedad o insatisfacción.	1	
	B4. Busca reducir estados indeseables de incertidumbre e incógnitas específicas.	1	

Dimensiones y características del instrumento.

Sobre la validación

La valoración del cuestionario se basa en cuatro categorías: Claridad, pertinencia, relevancia y suficiencia. En el caso de suficiencia, la valoración se realiza por dimensión, pues se valora si los ítem son suficientes o no para entregar información en relación con la correspondiente característica de la curiosidad. Para cada ítem se presenta una escala Likert que va de 1 a 4, además de considerar un apartado para valorar aspectos cualitativos desde su perspectiva como experto. La Tabla 2 presenta una síntesis de la categoría, la calificación y su respectivo indicador, para tener presente al momento de valorar cada pregunta.

Tabla 2.

Descripción de categorías y sus indicadores según la calificación.

Categoría	Calificación	Indicador
Claridad El ítem es claramente comprensible lingüísticamente.	1	El ítem no es comprensible o es altamente ambiguo . Su formulación impide una evaluación consistente y requiere una reescritura completa.
	2	El ítem presenta ambigüedades significativas en su redacción y requiere bastantes modificaciones en su estructura o terminología para ser preciso y evaluable.
	3	El ítem es comprensible en su mayoría , pero algún término o frase es vago o poco específico. Necesita una modificación puntual para lograr una redacción óptima.
	4	El ítem está formulado con precisión y sin ambigüedades . La conducta o actitud que se evalúa es clara y directamente observable para el evaluador.
Pertinencia El ítem planteado es pertinente en relación con la característica de la curiosidad que está midiendo.	1	El ítem planteado no es pertinente en relación con la característica de la curiosidad que mide.
	2	El ítem planteado es parcialmente pertinente en relación con la característica de la curiosidad que mide.
	3	El ítem planteado es pertinente en relación con la característica de la curiosidad que mide.
	4	El ítem planteado es totalmente pertinente en relación con la característica de la curiosidad que mide.
Relevancia El ítem planteado es relevante de acuerdo con el objetivo de la investigación, es decir, debe ser incluida.	1	El ítem planteado no es relevante y se puede eliminar .
	2	El ítem planteado tiene cierta relevancia , pero otro ítem puede estar midiendo algo similar.
	3	El ítem planteado es relativamente relevante .
	4	El ítem planteado es muy relevante y debe ser incluido .
Suficiencia Los ítem son suficientes en relación con la dimensión que se está midiendo.	1	Los ítems planteados no son suficientes.
	2	Los ítems planteados miden algunos aspectos de la dimensión.
	3	Los ítems planteados deben aumentar uno o dos más para medir la dimensión completamente.
	4	Los ítems planteados son suficientes.



Facultad de Educación
Universidad de Concepción

**INSTRUMENTO DE VALIDACIÓN: CURIOSIDAD EPISTÉMICA EN CONTEXTOS
DE RAZONAMIENTO INFERENCIAL INFORMAL (CE-RII)**

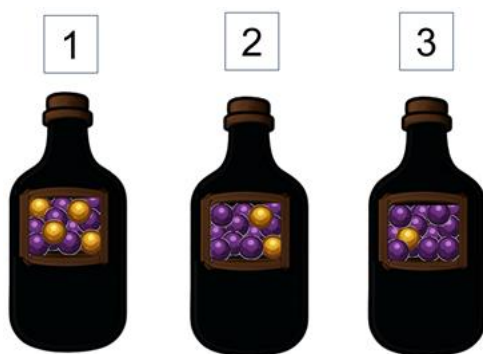
Antecedentes del validador

Nombre	
Filiación institucional	
Estudios de posgrado	
Línea de investigación principal	
Experiencia en relación con la Formación inicial docente en Matemáticas	

Evaluación de la Validez de constructo

A continuación, se presentan los ítems según cada dimensión y los criterios para la valoración a considerar en cada caso. Le solicitamos asignar la calificación de su valoración por criterio según cada pregunta y, al final de la dimensión, valorar a partir del criterio de suficiencia.

Observación: Previo a la entrega de este instrumento se compartirá con los estudiantes el siguiente contexto: “En una isla misteriosa, unos piratas aventureros encontraron un tesoro en tres botellas negras, tan bien selladas que no se pueden abrir (muestra las botellas). Sin embargo, el guardián de las botellas solo les permite llevarse una de ellas, por lo que deben elegir sabiamente. Afortunadamente, cada botella tiene una ventana, que permite observar parte de lo que lleva dentro. Cada botella carga algunas perlas de oro, muy valiosas y además, algunas perlas moradas, que son falsas”.



7.3 ANEXO 3: INSTRUMENTO PARA LA EVALUACIÓN DE ENTREVISTA COGNITIVA

INSTRUMENTO DE ENTREVISTA COGNITIVA - CURIOSIDAD EPISTÉMICA EN RAZONAMIENTO INFERENCIAL INFORMAL (CE-RII)

I. INFORMACIÓN GENERAL

Nombre del estudiante: _____

Edad: _____ Curso: _____

Fecha de entrevista: ____ / ____ / ____

Entrevistador: _____

Estimado participante:

Te damos la bienvenida a esta actividad. Tu participación es muy valiosa para nosotros.

El objetivo de esta entrevista es conocer tu opinión sobre un conjunto de preguntas y respuestas. Queremos saber si **las preguntas y respuestas son claras y fáciles de entender**. No es una evaluación, por lo que no hay respuestas correctas o incorrectas. Nos interesa conocer tu punto de vista, por lo que te pedimos seas muy honesto con nosotras al momento de responder.

Durante la entrevista, te pediremos que:

- Leas cada pregunta en voz alta o silenciosamente.
- Que nos cuentes qué piensas al leerla (qué significa para ti, si la entiendes o no).
- Nos digas si hay palabras que no entiendas.

Toda la información que nos compartas será un secreto, no le diremos a nadie. Solo la usaremos para entender qué mejorar en la encuesta.

Tu participación es voluntaria. Si en algún momento te quieres ir o quieres hacer una pausa, nos lo puedes decir.

Agradecemos mucho tu ayuda y tu tiempo.

II. RÚBRICA DE EVALUACIÓN

Instrucciones para el entrevistador: Escriba el número que corresponda al nivel que mejor refleje el nivel de cumplimiento de cada ítem, considerando los valores cualitativos de la escala likert e indicadores de los dominios a evaluar.

Etiqueta	Puntuación	Definición
Totalmente de acuerdo	5	El ítem cumple excelentemente. La evidencia es clara y fuerte en todos los aspectos (por ejemplo, los 4 indicadores son abordados de manera sobresaliente). El ítem no requiere modificaciones
De acuerdo	4	El ítem cumple bien con los indicadores. La evidencia es mayoritariamente positiva. El ítem requiere menores ajustes.

Neutral	3	El ítem cumple de manera aceptable pero no óptima con los indicadores. El ítem requiere ajustes importantes.
En desacuerdo	2	El ítem cumple deficientemente con los indicadores. El ítem requiere un ajuste estructural.
Totalmente en desacuerdo	1	El ítem no cumple en absoluto con los indicadores. El ítem debe ser rediseñado completamente.

Dominio a evaluar	Indicadores observables	Puntaje 1-5	Comentarios
Comprensión Semántica y Lingüística	(1) El niño no pregunta el significado de ninguna palabra. (2) El niño parafrasea la pregunta correctamente usando sinónimos adecuados. (3) Su expresión facial no denota confusión al leer el ítem. (4) No necesita que se le repita la pregunta de forma adicional.		
Comprensión de la Tarea y Estructura	(1) Comprende que debe elegir una sola opción de respuesta. (2) Sabe cómo marcar su respuesta (rodear, tachar, etc.) sin ayuda. (3) Entiende la instrucción general de la prueba (ej: "lee y elige la opción que más se parezca a ti"). (4) El formato de selección múltiple (a-b-c-d) le es familiar y manejable.		
Comprensión Conceptual (Validez de Constructo)	(1) Su explicación verbal del ítem se alinea con el constructo de curiosidad epistémica que se pretende medir (ej: "esto es sobre querer aprender cosas nuevas"). (2) La opción que eligió coincide con la justificación que da (validez interna de la respuesta). (3) No se distrae con elementos accesorios o irrelevantes del ítem.		

<p>Engagement y Pertinencia Percibida</p>	<p>(1) Mantiene la atención durante la lectura y respuesta del ítem. (2) Verbaliza que la situación presentada le es familiar o posible (ej: "eso me ha pasado en el cole"). (3) Muestra interés genuino por el contenido de la pregunta (ej: sonrío, se queda pensativo). (4) El tiempo que tarda en responder es apropiado (ni inmediato ni excesivamente largo por distracción).</p>		
---	--	--	--