



Facultad de educación

Programa Magíster en Evaluación Educacional

Percepción docente sobre fortalezas y debilidades del uso de ChatGPT en la creación de reactivos de comprensión lectora en cuarto básico

Tesis para optar al grado de Magíster en Evaluación Educacional

Patricio Tapia Leyton

Profesor guía: Dr. Miguel Alejandro Rodríguez Jara

**Valparaíso, Chile
2024**

RESUMEN

El presente estudio se desarrolla ante la necesidad de fortalecer los procesos de evaluación de la comprensión lectora en estudiantes de cuarto año básico, integrando herramientas tecnológicas emergentes al quehacer docente. En particular, se analiza el uso de ChatGPT como generador de ítems evaluativos en comprensión lectora, a partir de la percepción de docentes del sistema escolar chileno. Lo anterior, desde un enfoque cuantitativo, con diseño no experimental de tipo descriptivo, en donde se aplicó un cuestionario estructurado en escala Likert y previamente validado mediante un juicio de expertos. Los resultados evidencian fortalezas en la coherencia curricular, adaptabilidad y claridad de los ítems generados por la inteligencia artificial, así como debilidades asociadas a aspectos éticos, de validez pedagógica y equidad. Se concluye que ChatGPT puede constituir un recurso complementario y valioso en la construcción de instrumentos evaluativos, siempre que su implementación se acompañe de una reflexión crítica, fundamentos técnicos y resguardos éticos. Esta investigación entrega evidencia empírica relevante para orientar la toma de decisiones en el diseño y validación de instrumentos de evaluación en el contexto de la educación básica.

Palabras clave: *ChatGPT, Inteligencia artificial, comprensión lectora, reactivos evaluativos, percepción docente.*

ABSTRACT

This study examines teachers' perceptions regarding the technical quality and curricular alignment of reading comprehension items generated by ChatGPT for fourth-grade students in Chile. Using a quantitative, descriptive-exploratory design, a structured Likert-scale questionnaire was administered to 20 Language Arts teachers, focusing on clarity, perceived validity, curricular coherence, fairness, and practical applicability. Results revealed a high appreciation for the technical quality of the items, especially in terms of clarity and alignment with national learning objectives. However, concerns were raised about the equity and adaptability of the items to diverse classroom contexts. The findings suggest that artificial intelligence, when used with pedagogical guidance, can serve as a valuable tool in educational assessment, provided it is accompanied by critical review, teacher training, and ethical considerations.

Keywords: artificial intelligence, reading comprehension, educational assessment, teacher perception, ChatGPT.

Tabla de contenido

Índice de tablas y figuras	6
INTRODUCCIÓN	7
CAPÍTULO I: PROBLEMA DE INVESTIGACIÓN	9
1.1 Contexto y justificación del problema	9
CAPÍTULO II: MARCO CONCEPTUAL	13
2.1 Inteligencia artificial y ChatGPT en educación.....	13
2.2 Comprensión lectora como objeto de evaluación	14
2.3 Marco curricular y orientaciones ministeriales	15
2.4 Reactivos de evaluación: diseño y criterios de calidad	15
2.5 Percepción docente como evidencia evaluativa	16
CAPÍTULO III: DISEÑO METODOLÓGICO.	17
3.1 Paradigma de investigación.....	17
3.2 Tipo de estudio.....	17
3.3 Diseño	18
3.4 Unidad de análisis	18
3.5 Técnicas e instrumentos de recogida de la información.....	19
3.6 Técnicas de análisis.....	20
3.7 Consideraciones éticas.....	21
CAPÍTULO IV: RESULTADOS	22
4.1 Análisis descriptivo.....	22
4.2 Análisis mediante estadística implicativa.....	23
4.3 Estructuras Relacionales.....	25
4.3.1 Árbol de similaridad	25
4.3.2 Grafo implicativo	26
4.3.3 Árbol cohesitivo.....	27
4.4 Síntesis de resultados	28
CAPÍTULO V: DISCUSIÓN	30
5.1 Calidad técnica y alineación curricular de los ítems generados por ChatGPT	30
5.2 Aplicabilidad y consideraciones prácticas en el uso de los ítems.....	31

5.3 Inclusión, equidad y tensiones en la percepción docente	31
5.4 Reflexión crítica y contribuciones del estudio	32
CAPÍTULO VI: CONCLUSIONES	34
6.1 Conclusiones generales.....	34
6.2 Conclusiones según objetivos específicos.....	35
6.3 Proyecciones y líneas futuras de investigación	35
REFERENCIAS.	37

Índice de tablas y figuras

Tabla 1

Criterios de evaluación incluidos en el cuestionario aplicado a docentes -- -- 20

Figura 1

Árbol de Similitud entre los reactivos evaluados. ----- 37

Figura 2

Grafo Implicativo que representa las relaciones ----- 38
entre los ítems del cuestionario.

Figura 3

Árbol Cohesitivo que muestra la agrupación de ----- 39
percepciones entre los reactivos.

Figura 4

Árbol Cohesitivo que muestra la agrupación de ----- 40
percepciones entre los reactivos.

INTRODUCCIÓN

La comprensión lectora constituye una habilidad central para el aprendizaje en todas las áreas del conocimiento y un pilar fundamental para la participación en la sociedad del conocimiento (Solé, 1992; Cassany, 2006). Diversas evaluaciones internacionales, como PISA, y nacionales, como el SIMCE, han evidenciado preocupantes brechas en el desarrollo lector de los estudiantes chilenos, especialmente a partir del impacto de la pandemia de COVID-19, que profundizó las desigualdades educativas (OCDE, 2019; MINEDUC, 2023). Estas dificultades han impulsado la búsqueda de nuevas estrategias pedagógicas y evaluativas que contribuyan a mejorar la calidad del aprendizaje lector.

En este escenario, la inteligencia artificial (IA) se ha posicionado como una tecnología emergente con alto potencial para transformar los procesos de enseñanza y evaluación (Holmes et al., 2019). En particular, ChatGPT, desarrollado por OpenAI, ha demostrado capacidades para generar textos con coherencia semántica y contextual, lo que lo convierte en una herramienta interesante para la creación de reactivos de evaluación. No obstante, su uso en contextos escolares plantea desafíos vinculados a la validez, confiabilidad, equidad y pertinencia de los ítems generados, así como a la percepción y aceptación por parte del profesorado.

Considerando la creciente presencia de herramientas basadas en IA en el ámbito educativo, esta investigación se propone analizar las fortalezas y debilidades de los ítems de comprensión lectora generados por ChatGPT, desde la percepción de docentes de cuarto básico. Se espera así contribuir a la discusión sobre los usos pedagógicos de la IA en contextos reales de aula y aportar al campo de la evaluación educacional con evidencia empírica situada.

Este trabajo se estructura en siete capítulos. El primero aborda la problemática y justifica su relevancia en el contexto nacional. El segundo capítulo desarrolla el marco teórico, articulando conceptos clave como comprensión lectora, inteligencia artificial, reactivos evaluativos y percepción docente. El tercer capítulo expone el diseño metodológico, detallando el paradigma, el tipo de estudio, la muestra y el instrumento de recolección de

datos. En los capítulos siguientes se presentan los resultados, su análisis e interpretación, las conclusiones y proyecciones, destacando los aportes del estudio a la evaluación educacional en la era digital.

CAPÍTULO I: PROBLEMA DE INVESTIGACIÓN

1.1 Contexto y justificación del problema

La comprensión lectora constituye una competencia clave para el desarrollo de aprendizajes significativos, la formación del pensamiento crítico y la participación activa en la sociedad del conocimiento (Solé, 1992; Cassany, 2006). En el caso chileno, diversas mediciones nacionales e internacionales, como SIMCE y PISA, han evidenciado resultados insatisfactorios en esta área, especialmente en los primeros niveles del sistema escolar (MINEDUC, 2023; OCDE, 2019). Esta situación se ha visto agravada por los efectos de la pandemia de COVID-19, que profundizó las brechas de aprendizaje existentes, generando una necesidad urgente de implementar estrategias innovadoras que fortalezcan el desarrollo lector desde etapas tempranas.

En este escenario, las tecnologías emergentes, y en particular el avance de la inteligencia artificial (IA), han comenzado a posicionarse como herramientas con alto potencial para apoyar tanto los procesos de enseñanza como de evaluación (Hervás-Gómez et al., 2024). Entre estas, destaca ChatGPT, un modelo de lenguaje generativo desarrollado por OpenAI, que ha demostrado capacidad para producir textos coherentes, contextualizados y diversos. Esta funcionalidad abre la posibilidad de utilizar esta herramienta como apoyo en la elaboración de instrumentos evaluativos, especialmente reactivos de comprensión lectora. Sin embargo, su aplicación en contextos escolares plantea interrogantes fundamentales desde una perspectiva técnica y pedagógica: ¿Son válidos los ítems generados? ¿Se alinean adecuadamente con los objetivos curriculares? ¿Ofrecen condiciones de equidad para todos los estudiantes? ¿Cómo son percibidos por los docentes que eventualmente los aplicarían?

Si bien el uso de chatbots como ChatGPT en la educación aún está en fase inicial, su potencial ha sido ampliamente discutido. Kasneci et al. (2023) destacan su capacidad para personalizar el aprendizaje, proporcionar retroalimentación inmediata y apoyar la planificación de lecciones inclusivas. Asimismo, se ha señalado que esta herramienta puede facilitar la elaboración de actividades y evaluaciones alineadas con la taxonomía de Bloom, aunque con ciertas limitaciones en la precisión de sus respuestas (Kwan,

2024). Estudios como el de Küchemann et al. (2023) no encontraron diferencias sustanciales entre actividades diseñadas con ChatGPT y las extraídas de libros de texto, aunque sí reportaron errores en algunas respuestas generadas. De igual forma, Kwan (2024) detectó problemas en la asignación de puntajes y en la calidad de soluciones, observando una concentración en niveles bajos de la taxonomía de Bloom, como “recordar” y “comprender”, en detrimento de los niveles superiores como “evaluar” y “crear” (Sáiz, 2014; Gómez, 2014; Martínez y Gómez, 2018). También otros estudios han explorado el uso de ChatGPT en los ámbitos de la demanda cognitiva en tareas escolares de Historia (Cofré et al. ,2024) y Matemática (Puraivan et al., 2025).

El impacto de la IA en la educación también ha generado preocupaciones. Algunos docentes expresan inquietudes respecto de su posible uso para el fraude académico (Kelly, 2023). No obstante, otros autores argumentan que estas tecnologías deben entenderse como herramientas complementarias al quehacer docente (Ali et al., 2023; Eager y Brunton, 2023), capaces de aliviar la carga laboral del profesorado, especialmente en tareas administrativas o de generación de recursos (Gaete et al., 2017).

La limitada implementación de la IA en los procesos pedagógicos puede explicarse por factores como la preferencia de los docentes por materiales impresos o manipulativos, así como por la brecha generacional en el manejo de herramientas digitales (Ibáñez et al., 2017). A pesar de ello, su potencial en el área de comprensión lectora es considerable, especialmente por su capacidad de generar evaluaciones personalizadas y apoyar la mejora de los aprendizajes (Luckin et al., 2016). La efectividad de su integración, sin embargo, dependerá en gran medida de la percepción y aceptación de los docentes, quienes actúan como agentes clave en su implementación (Ertmer y Ottenbreit-Leftwich, 2010).

A pesar del creciente interés por incorporar IA en el ámbito educativo, aún existe escasa evidencia empírica sobre la calidad técnica, pedagógica y ética de los reactivos generados por modelos como ChatGPT. Esta falta de antecedentes representa una brecha relevante, considerando que la validez de un instrumento evaluativo no se limita a aspectos técnicos, sino que también debe considerar su pertinencia curricular y la

aceptación por parte de los actores educativos. Por ello, resulta pertinente indagar cómo los docentes valoran aspectos como la coherencia curricular, claridad, confiabilidad, aplicabilidad y justicia de los ítems generados mediante inteligencia artificial.

Este estudio se sitúa, por tanto, en la intersección entre comprensión lectora, evaluación educacional e inteligencia artificial, proponiendo un análisis exploratorio de las percepciones docentes respecto de ítems de comprensión lectora generados por ChatGPT para cuarto año básico. El propósito es identificar fortalezas y debilidades de dichos reactivos a partir de criterios de calidad técnica, adecuación pedagógica y equidad, con miras a entregar orientaciones informadas sobre su uso responsable en contextos escolares.

PREGUNTA DE INVESTIGACIÓN

¿Cuáles son las fortalezas y debilidades de los ítems de comprensión lectora generados por ChatGPT para cuarto básico, según la percepción docente, en relación con su calidad técnica, coherencia curricular y equidad?

OBJETIVO GENERAL.

Analizar las fortalezas y debilidades de los ítems de comprensión lectora generados por ChatGPT para cuarto básico desde la percepción docente, considerando criterios de calidad técnica, alineación curricular y equidad evaluativa.

OBJETIVOS ESPECÍFICOS.

1. Identificar las habilidades y niveles de comprensión lectora establecidos por el currículum nacional y referentes teóricos que deben considerarse en la elaboración de ítems evaluativos para cuarto básico.
2. Caracterizar la percepción de los docentes respecto a los ítems generados por ChatGPT en términos de su claridad, pertinencia, validez y adecuación pedagógica.
3. Determinar las principales fortalezas y debilidades atribuidas por los docentes a los ítems evaluativos generados por ChatGPT, considerando criterios de evaluación educacional como validez, confiabilidad y equidad.

CAPÍTULO II: MARCO CONCEPTUAL

Este capítulo expone los principales referentes conceptuales que sustentan la investigación, organizados en cinco ejes temáticos: la inteligencia artificial y ChatGPT en contextos educativos, la comprensión lectora como objeto de evaluación, los lineamientos curriculares vigentes en Chile, el diseño y análisis de reactivos en educación y, finalmente, la percepción docente como fuente de evidencia evaluativa. Cada uno de estos componentes contribuye a fundamentar el análisis de la calidad de los ítems generados por ChatGPT desde una perspectiva evaluativa.

2.1 Inteligencia artificial y ChatGPT en educación

La inteligencia artificial (IA) ha emergido como un conjunto de tecnologías que permiten a las máquinas simular funciones cognitivas humanas, como el aprendizaje, el razonamiento y el procesamiento del lenguaje natural (Russell y Norvig, 2020). En el ámbito educativo, la IA ha sido aplicada para personalizar la enseñanza, automatizar tareas administrativas y generar materiales didácticos adaptativos (Holmes et al., 2019).

Uno de los desarrollos más recientes es ChatGPT, un modelo generativo de lenguaje natural basado en transformadores, entrenado por OpenAI. Este modelo ha demostrado ser capaz de producir textos complejos, responder preguntas y generar ítems evaluativos a partir de consignas específicas. Su versatilidad ha llevado a explorar su potencial como apoyo en la creación de materiales de evaluación, incluyendo preguntas alineadas con marcos curriculares (Brown et al., 2020). Esta exploración se alinea con investigaciones recientes que han evaluado el potencial de ChatGPT para generar actividades educativas alineadas con la taxonomía de Bloom, destacando sus fortalezas en rapidez y adaptabilidad, aunque con ciertas limitaciones en niveles cognitivos superiores (Kwan, 2024; Küchemann et al., 2023; Puraivan et al., 2024).

Por otro lado, Hervás-Gómez et al. (2024) destacan que el uso pedagógico de IA exige una preparación crítica del profesorado, lo que condiciona su apropiación efectiva.

Además, se señala la importancia de evaluar los resultados que ofrecen estos sistemas en tareas escolares específicas.

El uso educativo de ChatGPT también plantea riesgos como la generación de contenido incorrecto, sesgos en los datos y limitaciones éticas asociadas al plagio, privacidad y fiabilidad (Floridi & Chiriatti, 2020; Crompton & Burke, 2024). Por ello, su aplicación en el diseño de ítems evaluativos requiere un análisis riguroso desde la perspectiva de la evaluación educacional.

2.2 Comprensión lectora como objeto de evaluación

La comprensión lectora es una habilidad compleja que involucra la decodificación, interpretación e integración del texto con el conocimiento previo del lector. Su desarrollo es fundamental para el aprendizaje en todas las áreas curriculares (Brigman et al., 2018) y se reconoce como una de las competencias básicas para el desempeño escolar y ciudadano. Además, es uno de los focos de evaluación de políticas públicas educacionales, como lo demuestran los estándares del SIMCE en el contexto chileno (MINEDUC, 2018).

Diversas investigaciones distinguen niveles de comprensión: literal, inferencial y crítica (Smith, 1989; Barrett, citado en Catalá et al., 2007). Estas categorías permiten clasificar ítems de evaluación según el tipo de procesamiento cognitivo requerido, lo que resulta clave al momento de analizar la calidad de los reactivos generados.

Barrett (1976) desarrolló una taxonomía específica para la comprensión lectora, la cual ha sido utilizada ampliamente en la construcción de instrumentos escolares. Dicha taxonomía establece niveles que van desde la recuperación de información explícita hasta la emisión de juicios críticos. Este marco ha sido adoptado por algunos estudios que buscaron automatizar la clasificación de ítems con ayuda de IA (Puraivan et al., 2024).

2.3 Marco curricular y orientaciones ministeriales

En el sistema educativo chileno, la comprensión lectora se encuentra normativamente regulada a través de las Bases Curriculares y los Programas de Estudio del Ministerio de Educación (MINEDUC, 2018). Para cuarto año básico, se establecen objetivos de aprendizaje que integran habilidades como inferir, comparar, interpretar lenguaje figurado, distinguir hechos de opiniones y emitir juicios críticos.

La evaluación de estas habilidades debe alinearse con los estándares de aprendizaje, lo cual exige que los reactivos utilizados tengan coherencia curricular, pertinencia cultural y claridad lingüística. La creación de ítems que cumplan estos criterios requiere, por tanto, un conocimiento profundo tanto del currículo como de los principios técnicos de la evaluación educacional.

2.4 Reactivos de evaluación: diseño y criterios de calidad

Los reactivos evaluativos constituyen instrumentos esenciales para recolectar evidencia del aprendizaje. Su calidad suele evaluarse considerando diversos criterios, tales como la validez, entendida como la capacidad del instrumento para medir lo que realmente se propone medir; la confiabilidad, referida a la consistencia de los resultados obtenidos; la equidad, que implica que el reactivo sea justo y accesible para todos los estudiantes; la claridad lingüística, que favorece la comprensión de las consignas; y la alineación con los objetivos de aprendizaje establecidos en el currículo (Nitko, 2001; Popham, 2003).

Según Covacevich (2014), al seleccionar o construir un instrumento, deben considerarse al menos cinco criterios: coherencia con el objetivo de evaluación, calidad técnica, viabilidad práctica, tipo de instrumento adecuado y consideraciones éticas. En el caso de ítems generados mediante IA, estos criterios deben ser cuidadosamente examinados.

Investigaciones recientes han comenzado a comparar la calidad técnica de reactivos humanos e inteligencia artificial, identificando ventajas en rapidez y diversidad, pero también debilidades en términos de precisión semántica, sesgos y ausencia de validación empírica (Kwan, 2024; Küchemann et al., 2023). Además, estudios como los de Martínez-Gómez (2018) y Sáiz (2014) advierten que muchas herramientas generadas por IA

tienden a centrarse en niveles bajos de la taxonomía de Bloom, descuidando habilidades superiores como analizar o evaluar.

2.5 Percepción docente como evidencia evaluativa

La percepción docente constituye una fuente relevante de información en estudios de validación de instrumentos. Si bien no sustituye los análisis estadísticos de validez o confiabilidad, ofrece evidencia indirecta sobre la aceptabilidad, claridad, relevancia y aplicabilidad de los ítems evaluados (Fitzpatrick et al., 2011).

En el contexto de esta investigación, se considera que los docentes, en tanto expertos en la implementación del currículo y conocedores de su grupo de estudiantes, pueden aportar juicios significativos sobre la adecuación técnica y pedagógica de los reactivos generados por ChatGPT, convirtiéndose así en actores clave en procesos de validación basada en contenido y pertinencia del instrumento (American Educational Research Association, 2014; Fitzpatrick et al., 2011). Esto se enmarca en una visión de evaluación educativa que incorpora tanto criterios técnicos como sociales en el análisis de instrumentos (Stake, 2004).

La presente investigación se sustenta en un marco teórico que articula elementos del currículo nacional, la teoría de la evaluación educativa, la comprensión lectora como competencia y el uso emergente de la inteligencia artificial en el ámbito educativo. Esta articulación permite abordar la problemática desde una perspectiva integral, centrada en analizar la percepción docente sobre la calidad de los ítems generados por ChatGPT, en función de criterios relevantes para la evaluación educativa en el siglo XXI.

CAPÍTULO III: DISEÑO METODOLÓGICO.

Este capítulo describe el enfoque metodológico que sustentó el desarrollo de la presente investigación. En coherencia con el objetivo general de comprender las percepciones docentes sobre el uso de ChatGPT en la generación de reactivos evaluativos de comprensión lectora, se detallan el paradigma adoptado, el tipo de estudio, el diseño, la muestra, las técnicas de recolección de datos y el procedimiento de análisis de la información.

3.1 Paradigma de investigación

La investigación se enmarca en el paradigma cuantitativo, el cual se fundamenta en la medición objetiva de variables, el uso de instrumentos estructurados y el análisis de datos mediante técnicas estadísticas (Creswell, 2014). Este paradigma permite identificar patrones, establecer relaciones y generalizar resultados en función de una muestra representativa.

Desde el enfoque evaluativo del programa, se adoptó una perspectiva criterial y exploratoria, que apunta a establecer juicios respecto a la calidad de los reactivos generados por inteligencia artificial. Dichos juicios se sustentan en criterios técnicos como la validez de contenido, la claridad, la coherencia curricular, la aplicabilidad y la equidad, tal como proponen Nitko (2001) y Covacevich et al. (2020). El interés central radica en describir y sistematizar las percepciones docentes en torno a estos criterios, concebidos como estándares básicos en la construcción y validación de instrumentos evaluativos.

3.2 Tipo de estudio

El estudio fue de tipo no experimental, ya que no se manipuló ninguna variable deliberadamente, sino que se observó el fenómeno tal como ocurre en su contexto natural (Hernández et al., 2014). Asimismo, se enmarca dentro del enfoque descriptivo-exploratorio, dado que busca caracterizar la percepción docente sobre los ítems generados por ChatGPT, sin establecer relaciones causales, pero sí permitiendo reconocer tendencias o elementos emergentes que orienten futuras investigaciones de mayor profundidad.

Este tipo de estudio resulta pertinente cuando se desea conocer cómo se manifiesta un fenómeno en determinado grupo poblacional, especialmente en áreas donde la evidencia empírica aún es incipiente, como es el caso del uso de IA generativa en contextos escolares.

3.3 Diseño

Se utilizó un diseño transversal, lo que implica que la recolección de datos se realizó en un solo momento del tiempo. Este enfoque permitió capturar una “fotografía” actual de las percepciones docentes en relación con la calidad y aplicabilidad de los reactivos generados mediante ChatGPT, en el marco de la evaluación de la comprensión lectora en cuarto año básico.

El diseño transversal, además de ser eficiente en términos de recursos y tiempo, resulta adecuado cuando el objetivo es describir una situación concreta, sin pretender establecer efectos longitudinales o transformaciones temporales (Bisquerra, 2009).

3.4 Unidad de análisis

La unidad de análisis estuvo constituida por docentes que se seleccionaron mediante un muestreo intencionado, considerando dos criterios fundamentales: experiencia profesional impartiendo la asignatura de Lenguaje, particularmente en cuarto básico o enseñanza básica, y disposición voluntaria para participar en el estudio.

En total, se incluyeron 20 docentes, provenientes de establecimientos con distintas dependencias administrativas (municipales, particulares subvencionados y particulares pagados), lo que permitió incorporar una diversidad contextual representativa del sistema escolar chileno.

La validez de esta muestra se sustenta en su carácter experto, dado que los docentes participantes poseen conocimiento directo del currículo, de las características del nivel evaluado y de las prácticas de evaluación aplicadas en el aula. Desde esta perspectiva, se reconoce su capacidad para emitir juicios informados y pertinentes sobre la calidad técnica, pedagógica y contextual de los ítems generados mediante inteligencia artificial.

3.5 Técnicas e instrumentos de recogida de la información

La información fue recolectada mediante un cuestionario tipo Likert estructurado, elaborado específicamente para esta investigación. Este tipo de instrumento presenta ventajas relevantes en estudios cuantitativos, ya que permite recoger una amplia gama de opiniones con base en escalas previamente definidas, lo que facilita el análisis e interpretación de los datos (García, 2003).

El diseño del instrumento consideró los criterios propuestos por Covacevich (2014) sobre la selección de instrumentos evaluativos. Dichos criterios fueron organizados en categorías clave que orientaron la formulación de preguntas dirigidas a evaluar aspectos fundamentales de los reactivos generados por ChatGPT, tales como:

- Claridad del lenguaje.
- Coherencia con el nivel curricular.
- Adecuación al tipo de comprensión lectora (literal, inferencial, crítica).
- Equidad y ausencia de sesgos.
- Validez técnica percibida.

Para asegurar la validez de contenido, el cuestionario fue sometido a juicio de expertos, quienes analizaron la claridad, pertinencia y relevancia de cada ítem. Asimismo, se llevó a cabo una prueba piloto con docentes de características similares a la muestra definitiva, lo que permitió realizar ajustes en la redacción y estructura de los ítems.

Como parte del diseño, se elaboró una tabla de especificaciones que organizó los ítems según cinco dimensiones fundamentales, alineadas con criterios técnicos de calidad en evaluación educativa. Esta estructura permitió garantizar la cobertura adecuada de los aspectos más relevantes para valorar los reactivos generados por inteligencia artificial. A continuación, se presenta la tabla correspondiente:

Tabla 1*Criterios de evaluación incluidos en el cuestionario aplicado a docentes*

Dimensión evaluada	Descripción de la dimensión	N° de Ejemplo de ítem (generado por ChatGPT)
Claridad lingüística	Evalúa si el enunciado del ítem es comprensible y libre de ambigüedad.	¿Qué hizo el personaje principal al final del cuento? (marcar alternativa)
Coherencia curricular	Mide el grado de alineación del ítem con los objetivos de aprendizaje de 4° básico.	¿Cuál es el objetivo de aprendizaje que mejor representa esta pregunta? (selección múltiple)
Tipo de comprensión lectora	Clasifica el tipo de habilidad lectora requerida (literal, inferencial, crítica).	¿Qué puede inferirse sobre la actitud del personaje frente al problema?
Equidad y sesgos	Considera si el ítem es culturalmente pertinente y accesible a todos los estudiantes.	¿Qué dificultades podrían tener estudiantes de contextos rurales con esta pregunta?
Validez técnica percibida	Refleja la percepción docente sobre si el ítem mide efectivamente lo que pretende medir.	¿Considera que esta pregunta permite evaluar comprensión inferencial de forma válida?

3.6 Técnicas de análisis

El análisis de los datos combinó un enfoque descriptivo y relacional, con el propósito de caracterizar las percepciones docentes y explorar patrones emergentes en sus respuestas. En una primera etapa, se aplicaron técnicas de estadística descriptiva, incluyendo el cálculo de frecuencias absolutas y relativas, promedios y desviaciones estándar, para identificar tendencias generales y variabilidad en las respuestas al cuestionario.

Complementariamente, se incorporó el uso de estadística implicativa (Gras et al., 2008), que permite identificar relaciones significativas entre categorías de respuesta y

representar gráficamente la estructura lógica subyacente de los datos categóricos. Esta técnica resultó especialmente pertinente debido al carácter exploratorio del estudio y al interés por descubrir agrupaciones y patrones de juicio docente que no serían visibles mediante análisis estadísticos tradicionales.

El uso de estadística implicativa permitió construir tres representaciones gráficas clave:

- Árbol de similitud, que muestra agrupaciones de ítems según similitudes en los patrones de respuesta de los participantes.
- Grafo implicativo, que identifica relaciones unidireccionales del tipo “si A, entonces B”, lo que facilita visualizar cómo ciertas percepciones están asociadas entre sí.
- Árbol cohesivo, que permite observar subgrupos de juicios docentes consistentes respecto a una o más dimensiones evaluadas.

Los análisis fueron realizados utilizando el software CHIC (Classification Hiérarchique, Implicative et Cohésive), una herramienta reconocida en el ámbito de las ciencias sociales y la educación por su capacidad para representar estructuras relacionales a partir de datos categóricos.

Esta combinación de análisis descriptivo y relacional permitió no solo caracterizar cuantitativamente las percepciones docentes, sino también profundizar en la comprensión de los vínculos entre criterios evaluativos, aportando una mirada más integral y visualmente estructurada sobre la valoración docente de los ítems generados por ChatGPT.

3.7 Consideraciones éticas

La investigación se desarrolló conforme a los principios éticos fundamentales de la investigación educativa, especialmente aquellos relacionados con la confidencialidad, el consentimiento informado y la participación voluntaria. Se garantizó en todo momento el anonimato de los participantes y se resguardó la privacidad de los datos recopilados, los cuales fueron utilizados exclusivamente con fines académicos.

Previo a la aplicación del instrumento, se entregó a los participantes una carta de consentimiento informado, donde se explicitaron los objetivos del estudio, el carácter voluntario de la participación y el derecho a retirarse en cualquier momento sin consecuencias. Asimismo, el instrumento y los procedimientos asociados fueron revisados y aprobados por el comité de ética correspondiente, asegurando el cumplimiento de las normativas vigentes.

CAPÍTULO IV: RESULTADOS

Este capítulo presenta los principales resultados obtenidos a partir del cuestionario aplicado a docentes de cuarto año básico, centrado en su percepción sobre los ítems de comprensión lectora generados por ChatGPT. El análisis se estructura en tres niveles complementarios: descriptivo, implicativo y relacional, integrando representaciones gráficas derivadas de la estadística implicativa, tales como el árbol de similitud, el grafo implicativo y el árbol cohesivo.

4.1 Análisis descriptivo

Los resultados descriptivos se organizaron en torno a las cuatro dimensiones evaluadas en el cuestionario: Coherencia curricular, Calidad del instrumento, Consideraciones prácticas, y Ética y justicia.

- **Coherencia curricular:** Los docentes evaluaron favorablemente la alineación de los ítems con los objetivos de aprendizaje del currículo de cuarto básico. Más del 75% de los encuestados calificó con 4 o 5 puntos (en una escala de 1 a 5) los ítems relacionados con la coherencia, destacando la correspondencia con los niveles de comprensión de la taxonomía de Barrett. Esto sugiere una percepción general de adecuación curricular de los ítems generados por la IA.
- **Calidad del instrumento:** Se observó una alta valoración en cuanto a la claridad del lenguaje y la validez percibida de los ítems. El 82% de los participantes consideró que los reactivos eran adecuados para evaluar la comprensión lectora. No obstante, algunos comentarios cualitativos recogidos en la etapa piloto

señalaron la necesidad de ajustes culturales en ciertos textos, particularmente en relación con referencias poco familiares para los estudiantes.

- **Consideraciones prácticas:** Esta dimensión también fue bien valorada. Un 78% de los docentes indicó que los ítems eran aplicables en un contexto real de aula, como un bloque de 90 minutos con un grupo de hasta 30 estudiantes. Sin embargo, surgieron observaciones sobre el esfuerzo adicional requerido para analizar los resultados de manera individualizada, lo cual podría limitar su implementación sin apoyo tecnológico o tiempo de planificación adicional.
- **Ética y justicia:** Aunque esta dimensión recibió en general valoraciones positivas, fue la que presentó mayor dispersión en las respuestas. Algunos docentes expresaron inquietudes respecto a la equidad cultural de los ítems y su adaptabilidad a estudiantes con necesidades educativas especiales. Esta variabilidad en las respuestas sugiere una necesidad de atención particular a la diversidad al momento de utilizar IA en contextos escolares.

4.2 Análisis mediante estadística implicativa

Para complementar el análisis descriptivo, se aplicó estadística implicativa, lo que permitió identificar relaciones significativas entre las respuestas entregadas por los docentes. Esta técnica ofrece una representación lógica y estructurada de los vínculos entre variables categóricas, destacando patrones que podrían pasar desapercibidos en un análisis tradicional.

Entre los hallazgos más relevantes se destacan:

- **12a → 12c (85%):** Se observó una fuerte implicación entre la percepción de aplicabilidad práctica de los textos A y C. Esto indica que, para una proporción significativa del profesorado, la utilidad en aula de distintos textos generados por ChatGPT se valora de manera coherente.
- **12a → 13a (84%):** Quienes consideran que los reactivos son aplicables en el aula, también tienden a percibir el lenguaje como accesible y no discriminatorio. Esta relación subraya el vínculo entre viabilidad técnica y justicia evaluativa.

- **13a → 13c (81%)**: Se evidenció una relación entre la percepción de lenguaje inclusivo y la adaptabilidad de los ítems a estudiantes con necesidades especiales. Sin embargo, la disminución leve en la fuerza del vínculo sugiere que la inclusividad lingüística no siempre garantiza adecuación plena a la diversidad educativa.
- **11a → 11b (75%)**: Se identificó una asociación entre la percepción del esfuerzo requerido para analizar los resultados y la valoración de la eficiencia administrativa del instrumento. Este hallazgo revela una posible tensión entre la aplicabilidad técnica y la carga laboral docente.

Existe una valoración general positiva respecto de los reactivos generados por ChatGPT, especialmente en dimensiones de coherencia y calidad. No obstante, las relaciones implicativas también evidencian áreas críticas, como la adaptabilidad a contextos diversos y el tiempo necesario para aplicar y revisar los instrumentos, que deben ser consideradas para su implementación efectiva en contextos escolares reales.

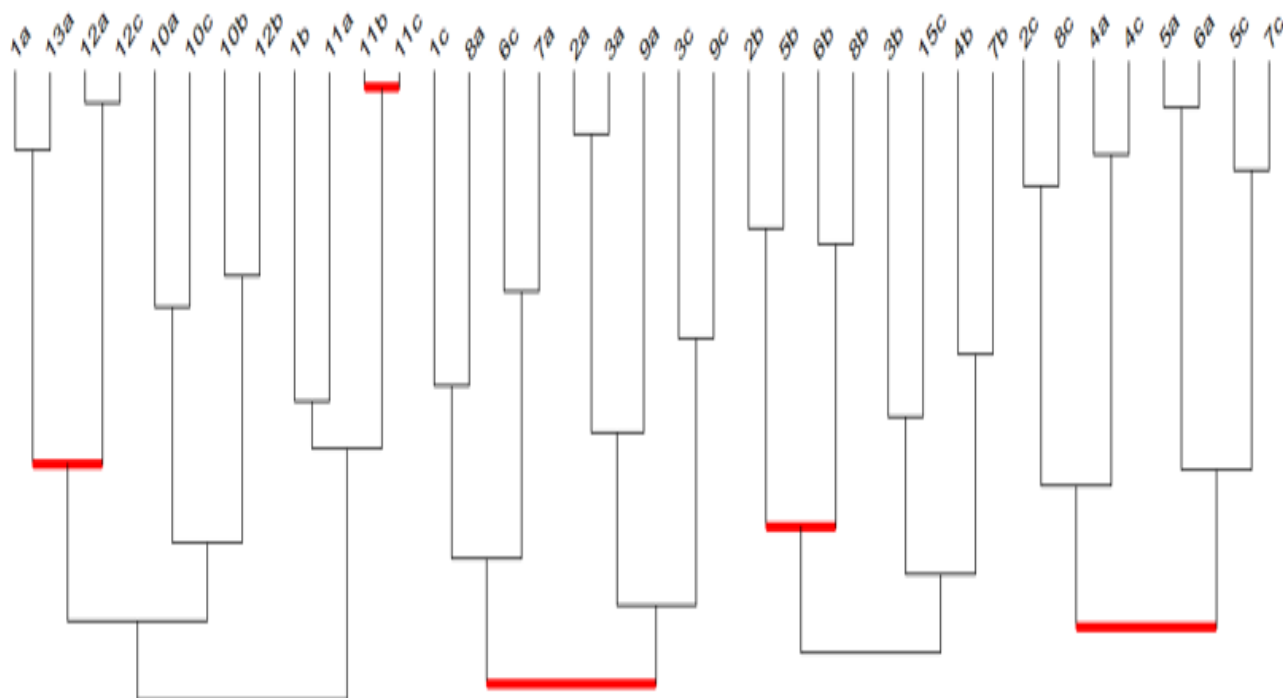
4.3 Estructuras Relacionales

4.3.1 Árbol de similaridad

Figura 1

Árbol de Similitud entre los reactivos evaluados.

Fuente: Elaboración propia a partir del análisis realizado con CHIC v7.0.



El árbol de similitud identificó patrones de respuesta altamente concordantes entre ítems asociados a la dimensión de practicidad, particularmente en los reactivos vinculados a los textos A y C (ítems 12a y 12c), con una similitud del 65,9%. Este agrupamiento sugiere una percepción consistente de aplicabilidad entre distintos textos generados por IA.

Asimismo, se evidenció una asociación entre ítems relacionados con alineación curricular y viabilidad de aplicación en aula (por ejemplo, ítems 1c y 10c). Esto indica que los docentes que perciben los ítems como coherentes con los objetivos de aprendizaje

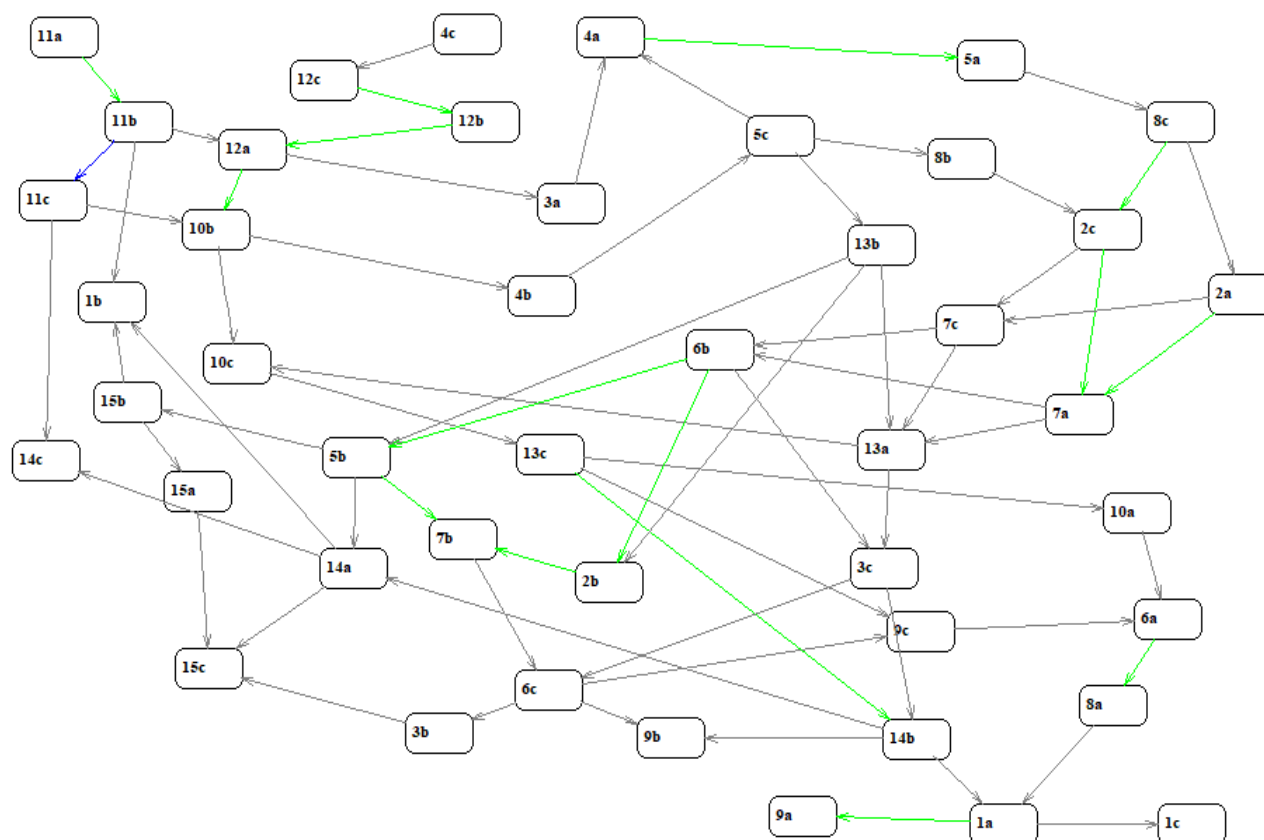
también tienden a considerarlos factibles de aplicar en condiciones reales de enseñanza, lo que refuerza la noción de coherencia técnica y funcional.

4.3.2 Grafo implicativo

Figura 2

Grafo Implicativo que representa las relaciones entre los ítems del cuestionario.

Fuente: Elaboración propia a partir del análisis realizado con CHIC v7.0.



El grafo implicativo permitió visualizar vínculos unidireccionales entre categorías, reforzando relaciones ya identificadas en el análisis descriptivo e implicativo. Entre los hallazgos destacados se encuentran:

- Conexiones sólidas entre claridad, aplicabilidad y accesibilidad lingüística, lo que refleja una valoración integrada de estos componentes por parte de los docentes.

Es decir, cuando un ítem es percibido como claro, también es considerado útil y accesible para el estudiantado.

- Presencia de tensiones en torno a la adaptabilidad y la equidad, especialmente en contextos de diversidad educativa. Aunque los ítems generados por ChatGPT son valorados positivamente en términos generales, persisten dudas sobre su adecuación para estudiantes con necesidades educativas especiales o provenientes de entornos culturales diversos.

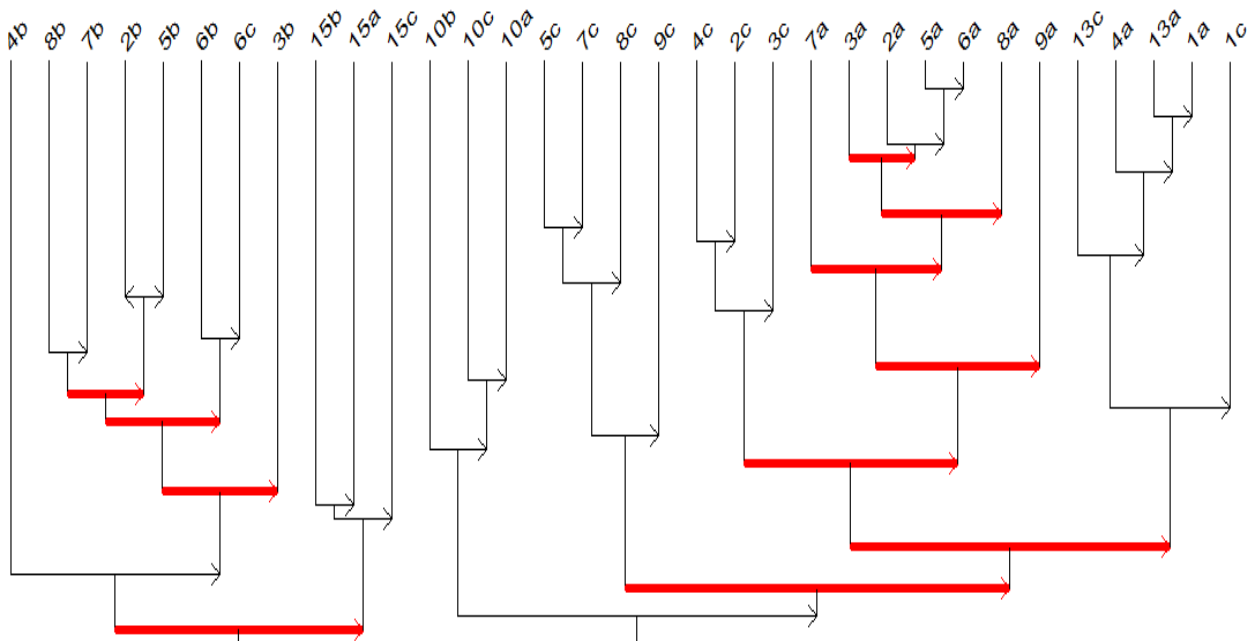
Estos resultados sugieren que los docentes operan con esquemas de valoración interrelacionados: quienes otorgan valor a los aspectos técnicos de los ítems tienden también a asignarles valor práctico y ético, dando cuenta de un juicio profesional complejo e integral.

4.3.3 Árbol cohesitivo

Figura 3

Árbol Cohesitivo que muestra la agrupación de percepciones entre los reactivos.

Fuente: Elaboración propia a partir del análisis realizado con CHIC v7.0.



El análisis mediante el árbol cohesitivo permitió observar estructuras jerárquicas de agrupación entre respuestas, evidenciando niveles diferenciados de consenso entre los docentes encuestados. Entre los hallazgos más relevantes se encuentran:

- Una alta cohesión inicial entre ítems asociados a calidad técnica y alineación curricular (por ejemplo, entre 5a y 6a, con una cohesión de 0.832). Este resultado indica un amplio acuerdo entre los docentes respecto a la adecuación pedagógica de los ítems generados.
- Una disminución progresiva en los niveles de cohesión al incorporar criterios asociados a accesibilidad, equidad y adaptabilidad, alcanzando un valor mínimo de cohesión de 0.659 en el nivel 20 del árbol. Esta tendencia refleja una mayor diversidad de opiniones cuando se abordan dimensiones éticas o contextuales, lo que evidencia la complejidad que implica implementar instrumentos evaluativos justos e inclusivos en escenarios educativos diversos.

4.4 Síntesis de resultados

En términos generales, los análisis realizados permiten establecer las siguientes conclusiones preliminares sobre la percepción docente respecto a los ítems de comprensión lectora generados por ChatGPT:

- Se observa una alta aceptación por parte de los docentes en lo relativo a la calidad técnica de los reactivos, así como a su coherencia con los objetivos curriculares establecidos para cuarto año básico.
- La practicidad de los ítems fue también bien valorada, particularmente en lo referente a su aplicación en condiciones reales de aula. No obstante, algunos docentes señalaron desafíos asociados al tiempo requerido para el análisis de resultados, lo que sugiere la necesidad de ajustes en este aspecto.
- Las mayores tensiones se identificaron en torno a los criterios de inclusión y justicia educativa, especialmente respecto a la adaptabilidad de los ítems a contextos de alta diversidad cultural y a estudiantes con necesidades educativas especiales.

Las representaciones gráficas derivadas del análisis implicativo (árbol de similaridad, grafo implicativo y árbol cohesitivo) refuerzan la consistencia del juicio docente en dimensiones técnicas y curriculares, mientras que evidencian una mayor dispersión de opiniones al incorporarse factores éticos y contextuales. Esto sugiere que, si bien los reactivos generados por IA presentan fortalezas claras en términos de calidad y alineación curricular, su implementación inclusiva y equitativa constituye un desafío aún por abordar.

CAPÍTULO V: DISCUSIÓN

Este capítulo tiene por propósito interpretar los principales hallazgos del estudio, integrándolos con el marco teórico y con los objetivos específicos planteados. La discusión se estructura en torno a tres ejes temáticos: (1) calidad técnica y alineación curricular de los ítems generados por ChatGPT, (2) percepción docente sobre su aplicabilidad y pertinencia en contextos reales de aula, y (3) tensiones asociadas a la equidad y justicia evaluativa. Finalmente, se ofrece una reflexión crítica que vincula los aportes del estudio con desafíos actuales en la evaluación educacional.

5.1 Calidad técnica y alineación curricular de los ítems generados por ChatGPT

Uno de los hallazgos más consistentes del estudio fue la valoración positiva que los docentes otorgaron a la calidad técnica de los ítems generados por ChatGPT. Esta percepción se reflejó en altos puntajes asignados a dimensiones como la claridad lingüística, la validez percibida y la coherencia con los objetivos de aprendizaje establecidos para cuarto básico.

Estos resultados coinciden con investigaciones recientes que han analizado el potencial de los modelos de lenguaje generativos en la creación de instrumentos evaluativos (Kwan, 2024; Küchemann et al., 2023). Se confirma, así, que ChatGPT es capaz de generar preguntas que se ajustan tanto a marcos curriculares como a estructuras cognitivas reconocidas, tales como la taxonomía de Barrett (Catalá et al., 2007). Desde una perspectiva evaluativa, los docentes reconocieron la validez de contenido de los ítems, entendida como la correspondencia entre el objetivo evaluativo y lo efectivamente medido (Nitko, 2001; AERA, 2014).

El uso del juicio experto como estrategia de validación en este estudio resultó fundamental para sustentar la aceptabilidad técnica de los reactivos, y se enmarca en una tradición evaluativa que reconoce el valor del conocimiento profesional docente como evidencia legítima (Fitzpatrick et al., 2011).

5.2 Aplicabilidad y consideraciones prácticas en el uso de los ítems

El segundo eje de análisis abordó la viabilidad de implementar los ítems generados por ChatGPT en situaciones reales de enseñanza. La mayoría de los docentes (78%) consideró que estos reactivos podían aplicarse adecuadamente en condiciones habituales de aula (por ejemplo, sesiones de 90 minutos con cursos de hasta 30 estudiantes), lo que refuerza la hipótesis de que la IA puede constituir un recurso útil para apoyar la labor docente, especialmente en el ámbito de la evaluación.

Este hallazgo adquiere particular relevancia en el contexto educativo chileno, donde las exigencias administrativas y curriculares son elevadas. En este sentido, autores como Holmes et al. (2019) han sostenido que el diseño asistido por IA puede optimizar recursos y reducir la carga docente, siempre que se mantenga una supervisión crítica de los productos generados.

No obstante, emergieron advertencias relacionadas con el tiempo requerido para analizar los resultados de las evaluaciones. Las relaciones implicativas encontradas entre la percepción de esfuerzo y aplicabilidad sugieren que, si bien los ítems son técnicamente sólidos, su uso efectivo implica una preparación previa y una estrategia de retroalimentación bien definida. Esto evidencia una tensión entre eficiencia técnica y viabilidad operativa, que podría ser resuelta mediante formación docente específica en evaluación asistida por tecnologías.

5.3 Inclusión, equidad y tensiones en la percepción docente

Un hallazgo relevante del estudio fue la mayor dispersión de respuestas en torno a los criterios de inclusión y justicia educativa. Aunque la mayoría de los docentes evaluó positivamente la dimensión ética, algunos manifestaron dudas respecto a la adaptabilidad de los ítems a contextos de alta diversidad cultural o con presencia de estudiantes con necesidades educativas especiales.

Este resultado se vincula con la literatura que advierte sobre los sesgos potenciales en el uso de tecnologías de inteligencia artificial, especialmente en lo que respecta a lenguaje, cultura y accesibilidad (Floridi & Chiriatti, 2020). En el presente estudio, la

estadística implicativa mostró vínculos más débiles entre claridad, justicia y adaptabilidad, lo que sugiere que estos aspectos son percibidos como menos consistentes o más problemáticos que otros.

Frente a ello, se hace evidente la necesidad de procesos de revisión crítica y contextualización cultural de los ítems generados por IA. Tal como lo plantean Stake (2004) y Covacevich et al. (2020), una evaluación justa no puede limitarse a criterios técnicos, sino que debe incorporar dimensiones éticas, socioculturales y de equidad en su diseño y aplicación.

5.4 Reflexión crítica y contribuciones del estudio

Desde un enfoque de evaluación educacional integral, este estudio aporta evidencia empírica sobre la factibilidad técnica y la percepción docente respecto al uso de inteligencia artificial en la elaboración de reactivos de comprensión lectora. Si bien los resultados son prometedores, también dejan en evidencia las condiciones necesarias para una implementación responsable de estas tecnologías en el sistema escolar.

En concreto, el estudio sugiere que el uso de ChatGPT en contextos educativos debe considerar:

- Procesos sistemáticos de validación empírica, más allá del juicio experto.
- Revisión cultural y lingüística situada, que garantice la pertinencia y accesibilidad de los ítems.
- Formación docente específica, centrada en el diseño y uso crítico de instrumentos evaluativos.
- Análisis ético continuo, orientado a evitar sesgos y garantizar la transparencia y privacidad en el uso de herramientas digitales.

El principal valor de este trabajo radica en haber situado a los docentes como agentes evaluadores, reconociendo su experiencia profesional como una forma válida de evidencia en el análisis de instrumentos. Esta perspectiva responde al llamado de Fitzpatrick et al. (2011) a integrar enfoques técnicos, sociales y éticos en los procesos

evaluativos contemporáneos, especialmente cuando se incorporan tecnologías emergentes como la inteligencia artificial.

CAPÍTULO VI: CONCLUSIONES

Este capítulo presenta las principales conclusiones del estudio y propone líneas de proyección futura en el ámbito de la evaluación educacional asistida por inteligencia artificial. Las conclusiones se organizan en función de los objetivos específicos de la investigación y se acompañan de reflexiones que invitan a profundizar en nuevas áreas de desarrollo, considerando oportunidades, desafíos pedagógicos y exigencias éticas.

6.1 Conclusiones generales

La presente investigación analizó, desde la percepción docente, las fortalezas y debilidades de los ítems de comprensión lectora generados por ChatGPT para cuarto año de educación básica. A partir del análisis cuantitativo, implicativo y gráfico-relacional, se destacan las siguientes conclusiones generales:

1. **Alta valoración técnica:** Los ítems generados por ChatGPT fueron altamente valorados por los docentes en cuanto a claridad, validez técnica y alineación curricular. Estos resultados evidencian que la inteligencia artificial, debidamente utilizada, puede transformarse en una herramienta significativa de apoyo a los procesos evaluativos, especialmente en lectura comprensiva.
2. **Consistencia en el juicio profesional docente:** El análisis implicativo reveló patrones lógicos entre las dimensiones evaluadas. Docentes que valoraron positivamente la coherencia curricular tendieron también a valorar la claridad y la aplicabilidad. Esto refuerza la validez del juicio docente como fuente legítima de evaluación técnica y pedagógica.
3. **Tensiones en torno a la equidad:** A pesar de los resultados favorables en aspectos técnicos, se evidenciaron preocupaciones respecto a la adaptabilidad de los ítems a contextos diversos y a estudiantes con necesidades educativas especiales. Estas tensiones advierten sobre la necesidad de procesos de revisión ética y contextual antes de la aplicación de ítems generados por IA en entornos reales.

6.2 Conclusiones según objetivos específicos

- **Objetivo 1: Identificar las habilidades y niveles de comprensión lectora establecidos por el currículo nacional.**

El estudio permitió mapear los Objetivos de Aprendizaje del currículum chileno vinculados a la comprensión lectora, confirmando que los ítems generados abordan adecuadamente los tres niveles propuestos por la taxonomía de Barrett: comprensión literal, inferencial y crítica. Esta adecuación fue reconocida por los docentes participantes.

- **Objetivo 2: Caracterizar la percepción docente sobre los ítems generados por ChatGPT.**

La percepción fue mayoritariamente favorable. Las dimensiones más valoradas fueron la claridad lingüística, la coherencia curricular y la viabilidad de aplicación en contextos reales de aula. Los docentes destacaron la potencial utilidad de los reactivos como recursos de apoyo evaluativo, sin necesidad de ajustes mayores.

- **Objetivo 3: Determinar fortalezas y debilidades de los ítems según criterios evaluativos.**

Entre las principales fortalezas se encuentran la claridad en la formulación, la alineación con el currículum nacional, la economía de diseño y la posibilidad de reutilización. Las debilidades identificadas se relacionan con la adaptabilidad en contextos culturalmente diversos, el riesgo de sesgos implícitos y la sobrecarga que puede suponer el análisis de resultados sin herramientas de apoyo.

6.3 Proyecciones y líneas futuras de investigación

Los hallazgos obtenidos permiten vislumbrar diversas rutas para el desarrollo futuro de la evaluación asistida por inteligencia artificial:

1. **Profundizar la validación empírica de los reactivos generados por IA,** incorporando análisis de desempeño estudiantil, revisión por pares y comparación con ítems de evaluaciones estandarizadas.

2. **Diseñar herramientas complementarias**, como analizadores automáticos de resultados, rúbricas digitales o plataformas de retroalimentación inmediata, que alivien la carga docente y fortalezcan la eficiencia del proceso evaluativo.
3. **Integrar criterios de equidad e inclusión** como parámetros explícitos en los modelos de generación automática de ítems, promoviendo el Diseño Universal para el Aprendizaje (DUA) y asegurando accesibilidad para todos los estudiantes.
4. **Fortalecer la formación docente inicial y continua** en el uso crítico, ético y pedagógico de herramientas basadas en IA, promoviendo una apropiación reflexiva de estas tecnologías.

Esta investigación ha demostrado que la inteligencia artificial, utilizada de manera ética, situada y pedagógicamente orientada, puede constituirse en una herramienta potente y legítima para apoyar la evaluación educacional. No obstante, su implementación efectiva requiere del juicio profesional docente, del acompañamiento institucional y del desarrollo de una cultura evaluativa crítica. El desafío es avanzar hacia un uso de la IA que no solo busque eficiencia, sino que también contribuya a una educación más justa, inclusiva y transformadora.

REFERENCIAS.

- Abdel-Messih, M. S., & Boulos, M. N. K. (2023). ChatGPT in Clinical Toxicology. *JMIR Medical Education*, 9, e46876. <https://doi.org/10.2196/46876>
- Ali, J. K. M., Shamsan, M. A. A., Hezam, T. A., & Mohammed, A. A. Q. (2023). Impact of ChatGPT on Learning Motivation: *Journal Of English Studies In Arabia Felix*, 2(1), 41-49. <https://doi.org/10.56540/jesaf.v2i1.51>
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- Agencia de Calidad de la Educación. (2023). Principales conclusiones de los resultados del Simce 2023. Recuperado de <https://resultadossimce.agenciaeducacion.cl/>
- Aguirre Rueda, J. A. (2015). El manual escolar como objeto de investigación en historia de la educación: Apuntes al texto de geografía. *Historia y Sociedad*, 28, 247–272. <https://doi.org/10.15446/hys.n28.48129>
- Alliende, F., y Condemarín, M. (1986). *La lectura: Teoría, evaluación y desarrollo*. Editorial Andrés Bello.
- Babbie, E. (2015). *The practice of social research Cengage Learning*
- Baidoo-Anu, D., & Ansah, L. O. (2023). Education in the Era of Generative Artificial Intelligence (AI): Understanding the Potential Benefits of ChatGPT in Promoting Teaching and Learning. *Journal Of AI*, 7(1), 52-62. <https://doi.org/10.61969/jai.1337500>
- Bazeley, P., & Jackson, K. (2013). *Qualitative Data Analysis with NVivo*. Sage Publications.
- Bisquerra, R. (2009). *Metodología de la investigación educativa* (2.ª ed.). La Muralla.

- Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of Educational Objectives: The Classification of Educational Goals*. Handbook I: Cognitive Domain. David McKay Company.
- Bolaño-García, M., & Duarte-Acosta, N. (2024). Una revisión sistemática del uso de la inteligencia artificial en la educación. *Revista Colombiana de Cirugía*, 39(1), 51–63. <https://doi.org/10.30944/20117582.2365>
- Bowyer-Crane, C., Snowling, M. J., Duff, F. J., Fieldsend, E., Carroll, J. M., Miles, J., Götz, K., & Hulme, C. (2008). Improving early language and literacy skills: differential effects of an oral language versus a phonology with reading intervention. *Journal Of Child Psychology And Psychiatry And Allied Disciplines*, 49(4), 422-432. <https://doi.org/10.1111/j.1469-7610.2007.01849.x>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101.
- Brigman, G. A., Webb, L. D., & Campbell, C. (2007). Building Skills for School Success: Improving the Academic and Social Competence of Students. *Professional School Counseling*, 10(3), 279-288. <https://doi.org/10.1177/2156759x070100031>
- Briones, G. (2002). *Metodología de la investigación cuantitativa en las ciencias sociales*. Instituto Colombiano para el Fomento de la Educación Superior (ICFES).
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*. <https://arxiv.org/abs/2005.14165>
- Byram, M., Nichols, A., & Stevens, D. (2013). *Developing Intercultural Competence in Practice*. Multilingual Matters.
- Cain, K. (2010). *Reading Development and Difficulties*. Editorial John Wiley & Sons.

- Camarillo Salazar, B. F., Silva Maceda, G., & Romero Contreras, S. (2021). El Modelo Simple de Lectura en la identificación de dificultades lectoras en educación primaria. *Estudios Pedagógicos*, 47(3), 343–357.
- Cassany, D., Luna, M., & Sanz, G. (2005). *Enseñar lengua*. Editorial Graó.
- Castellanos-Páez, V., Abello-Correa, R., Gutiérrez-Romero, M., Ochoa-Angrino, S., Rojas, T., Taborda-Osorio, H. (2022). Impacto de la pandemia en el aprendizaje: reflexiones desde la psicología educativa. *Praxis & Saber*, 13(34), e14532. <https://doi.org/10.19053/22160159.v13.n34.2022.14532>
- Catalá, G., Catalá M., Molina E., Monclús R. (2007). *Evaluación de la comprensión lectora: Pruebas ACL (1°-6° de primaria)*.
- Cofré, C., Puraivan, E., Huencho-Iturra, K., Macarena Astudillo-Vásquez, & M. Dolores Díaz-Noguera. (2024, November 6). *Using Artificial Intelligence to Assess the Level of Cognitive Complexity Involved in Didactic Tasks. A case study on Historical Thinking*. <https://doi.org/10.36006/09643-1-13>
- Condemarín, M., & Medina, A. (1999). *Taller de lenguaje II: Un programa integrado de desarrollo de las competencias lingüísticas y comunicativas de los alumnos de segundo ciclo básico*.
- Consejo Nacional de la Cultura y las Artes & Centro de Microdatos de la Universidad de Chile. (2011). *Estudio sobre el comportamiento lector a nivel nacional: Informe final 2011*. Recuperado de plandelectura.gob.cl/recursos/estudio-sobre-comportamiento-lector-a-nivel-nacional-2011/
- Constenla, J. A., Vera, A. J., Jara, P. A. (2022). Actitudes y capacidades de los docentes frente a la innovación educativa. La mirada de los estudiantes. *Pensamiento Educativo: Revista de Investigación Educativa Latinoamericana*, 59(1), 1–15. <https://doi.org/10.7764/pel.59.1.2022.7>

- Couturier, R. (2008). CHIC: Cohesive Hierarchical Implicative Classification. *Studies in Computational Intelligence*, 41–53. https://doi.org/10.1007/978-3-540-78983-3_2
- Crompton, H., & Burke, D. (2024). The Educational Affordances and Challenges of ChatGPT: State of the Field. *TechTrends*, 68(2), 380-392. <https://doi.org/10.1007/s11528-024-00939-0>
- Covacevich, C. (2014). Cómo seleccionar un instrumento para evaluar aprendizajes estudiantiles. *Banco Interamericano de Desarrollo*.
- Designano N. (2006). *Didáctica de las Ciencias del Lenguaje: Enseñar ciencias del lenguaje*. Editorial Homosapiens Ediciones.
- Díaz-Díaz, M., Echegoyen, Y., Martín-Ezpeleta, A. (2022). Del lector ideal al mediador competente. *Ocnos. Revista de Estudios Sobre Lectura*, 21(2). https://doi.org/10.18239/ocnos_2022.21.2.2967
- Eager, B., & Brunton, R. (2023). Prompting Higher Education Towards AI-Augmented Teaching and Learning Practice. *Journal Of University Teaching And Learning Practice*, 20(5). <https://doi.org/10.53761/1.20.5.02>
- Ertmer, P. A., & Ottenbreit-Leftwich, A. T. (2010). Teacher technology change: How knowledge, confidence, beliefs, and culture intersect. *Journal of Research on Technology in Education*, 42(3), 255–284. <https://doi.org/10.1080/15391523.2010.10782551>
- Escobar-Pérez, J., & Cuervo-Martínez, Á. (2008). *Validez de contenido y juicio de expertos: una aproximación a su utilización*. Avances en Medición.
- Flavell, J. H. (1997). El desarrollo del conocimiento de los niños sobre el habla interior. *Desarrollo del niño*, 68(1), 39–47.

Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2011). *Program evaluation: Alternative approaches and practical guidelines* (4th ed.). Pearson Education.

Floridi, L., & Chiriatti, M. (2020). GPT-3: Its nature, scope, limits, and consequences. *Minds and Machines*, 30(4), 681–694. <https://doi.org/10.1007/s11023-020-09548-1>

García, J. (2003). *Métodos en investigación en educación*. (V. II) Madrid: UNED.

García Madruga, J. (2006). *Lectura y conocimiento*. Ediciones Paidós Ibérica.

Garner, R. (1987). *Metacognición y comprensión lectora*.

Gaete Silva, A., Castro Navarrete, M., Pino Conejeros, F., & Mansilla Devia, D. (2017). Abandono de la profesión docente en Chile: Motivos para irse del aula y condiciones para volver. *Estudios Pedagógicos (Valdivia)*, 43(1), 123–138. <https://doi.org/10.4067/s0718-07052017000100008>

Gholami, M., Shirazi, T., Nikookar, N., & Rahimi, A. (2016). Comparison of effectiveness of phonetic intervention, holistic intervention, and mixed intervention on the rate and accuracy of reading skills of children with reading difficulty in grade 2. *The IOAE Journal*, 7(1), 270–276.

Giannos, P., & Delardas, O. (2023). Performance of ChatGPT on UK Standardized Admission Tests: Insights From the BMAT, TMUA, LNAT, and TSA Examinations. *JMIR Medical Education*, 9, e47737. <https://doi.org/10.2196/47737>

Gómez, C. (2014). Pensamiento histórico y contenidos disciplinares en los libros de texto. Un análisis exploratorio de la Edad Moderna en 2º de la ESO. *ENSAYOS. Revista de la Facultad de Educación de Albacete*, 29(1), 131-158. <https://revista.uclm.es/index.php/ensayos/article/view/498>

- Gómez, C. & Miralles, P. (2016). Développement et évaluation des compétences historiques dans les manuels scolaires. Une étude comparative France-Espagne. *Spirale – Revue de recherches en education*, 2(58), 53-66. <https://www.cairn.info/revue-spirale-revue-de-recherches-en-education-2016-2-page-53.htm>
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
- Gras, R., Suzuki, E., Guillet, F., & Spagnolo, F. (2008). *Statistical implicative analysis: theory and applications*. Springer.
- Grellet, F. (1981). *Developing reading skills*. Cambridge University Press.
- Haladyna, T. M. (2004). *Developing and validating multiple-choice test items*. Routledge.
- Hashem, R., Ali, N., El Zein, F., Fidalgo, P. & Abu Khurma, O. (2024). AI to the rescue: Exploring the potential of ChatGPT as a teacher ally for workload relief and burnout prevention. *Research and Practice in Technology Enhanced Learning*, 19. <https://doi.org/10.58459/rptel.2024.19023>
- Haviland, S. E., & Clark, H. H. (1974). What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Behavior*, 13, 512–521.
- Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, P. (2014). *Metodología de la investigación* (6.a ed.). McGraw-Hill.
- Hernández-Valle, I., & Jiménez, J. E. (2001). Conciencia fonémica y retraso lector: ¿Es determinante la edad en la eficacia de la intervención? *Infancia y Aprendizaje*, 24(3), 379–395. <https://doi.org/10.1174/021037001316949284>
- Hervás-Gómez, C., Díaz-Noguera, M. D., Puraivan, E., Astudillo-Vásquez, M., & Cofré-Morales, C. (2024). *Introduction to artificial intelligence in education*. Springer Nature.

- Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial Intelligence in Education: Promises and Implications for Teaching and Learning*. Center for Curriculum Redesign.
<https://curriculumredesign.org/wp-content/uploads/AI-in-Education-CCR-2019.pdf>
- Ibáñez, R., Moncada, F., Cornejo, F., & Arriaza, V. (2017). Los Géneros del Conocimiento en Textos Escolares de educación primaria. *Calidoscópico*, 15(3), 462–476. DOI: 10.4013/cld.2017.153.06.
- Jones, E., Larsen, R., Sudweeks, R. R., Young, K. R., & Gibb, G. S. (2018). Evaluating Paraeducator-led Reading Interventions in Elementary School: A Multi-Cutoff Regression-Discontinuity Analysis. *Journal Of Research On Educational Effectiveness*, 11(4), 507-534. <https://doi.org/10.1080/19345747.2018.1481164>
- Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., Kasneci, G. (2023). *ChatGPT for good? On opportunities and challenges of large language models for education*. Center for Open Science.
<http://dx.doi.org/10.35542/osf.io/5er8f>
- Küchemann, S., Steinert, S., Revenga, N., Matthias Schweinberger, Yavuz Dinc, Avila, K. E., & Jochen Kühn. (2023). Can ChatGPT support prospective teachers in physics task development? *Physical review physics education research*, 19(2).
<https://doi.org/10.1103/physrevphyseducres.19.020128>
- Kwan, C. (2024). *Exploring ChatGPT-Generated Assessment Scripts of Probability and Engineering Statistics from Bloom's Taxonomy* [Conference]. International Conference on Technology in Education, Hong Kong, China.
https://doi.org/10.1007/978-981-99-8255-4_24
- Levin, O., & Baratz, L. (2019). Reading in order to teach reading: Processive literacy as a model for overcoming difficulties. *L1 Educational Studies in Language and Literature*, 19, Running Issue(Running Issue), 1–24. <https://doi.org/10.17239/l1esll-2019.19.01.08>

- Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence Unleashed: An argument for AI in Education*. Pearson.
<https://www.pearson.com/content/dam/one-dot-com/one-dot-com/global/Files/news/news-announcements/2016/intelligence-unleashed.pdf>
- Russell, S., & Norvig, P. (2020). *Artificial Intelligence: A Modern Approach* (4th ed.). Pearson.
- Martínez, M. & Gómez, C. (2018). Cognitive level and thinking historically competencies in history textbooks from Spain and England: a comparative study. *Revista de Educación*, (379), 136-159. <https://doi.org/10.4438/1988-592X-RE-2017-379-364>
- Martínez-Serrano, M. C. (2019). Percepción de la Integración y uso de las Tecnologías de la Información y la Comunicación (TIC). Estudio de Profesores y Estudiantes de Educación Primaria. *CIT Informacion Tecnologica*, 30(1), 237–246.
<https://doi.org/10.4067/s0718-07642019000100237>
- Ministerio de Educación. (2023). *Plan de Desarrollo Lector y Orientaciones Curriculares*. Santiago, Chile.
- Ministerio de Educación de Chile. (2019). *Estándares de Aprendizaje: Lectura 4º Básico*. Unidad de Currículum y Evaluación.
- Ministerio de Educación de Chile. (2018). *Lenguaje y Comunicación: Programa de Estudio Cuarto Año Básico*. Unidad de Currículum y Evaluación.
- Norman-Acevedo, E. (2023). La inteligencia artificial en la educación: Una herramienta valiosa para los tutores virtuales universitarios y profesores universitarios. *Panorama*, 17(32), 1–11.
<https://doi.org/10.15765/pnrm.v17i32.3681>
- OECD (2019), *PISA 2018 Results (Volume I): What Students Know and Can Do*, PISA, OECD Publishing, Paris, <https://doi.org/10.1787/5f07c754-en>.

Orús, P.; Zamora, L. y Gregori, P. (2009). *Teoría y aplicaciones del análisis estadístico implicative: primera aproximación en lengua hispana*. Pilar Orús, Larisa Zamora, Pablo Gregori (Editores). Castellón: Universitat Jaume I, Departamento de Matemáticas; Santiago de Cuba: Facultad de Matemática y Computación. Universidad de Oriente de Santiago de Cuba, 2009. 516 p. ISBN 978-84- 692-3925-4.

Padilla, R. D. M. (2019). La llegada de la inteligencia artificial a la educación. *Revista de Investigación En Tecnologías de la Información*, 7(14), 260-270. <https://doi.org/10.36825/riti.07.14.022>

Panaqué, C. R., & Castañón, C. B. (2024). La inteligencia artificial en la educación del siglo XXI: avances, desafíos y oportunidades. *Educación/Educación*, 33(64), 5-7. <https://doi.org/10.18800/educacion.202401.p001>

Perines, H. (2018). La investigación educativa: percepción de docentes en ejercicio, docentes en formación y agentes educativos. *Perspectiva educacional*, 57(2). <https://doi.org/10.4151/07189729-vol.57-iss.2-art.649>

Popham, W. J. (2002). *Classroom Assessment: What Teachers Need to Know*. Allyn and Bacon.

Puraivan, E., Cofré-Morales, C., Rodríguez, M., Lasnibat-Godoy, T., Tapia Sobarzo, J. E., Hervás-Gómez, C., & Díaz-Noguera, M. D. (2025). ChatGPT and Semiotic Representation Theory: Innovating the Measurement of Cognitive Demand in Mathematical Tasks. *Distributed Computing and Artificial Intelligence* https://doi.org/10.1007/978-3-031-82073-1_23

Puraivan, E., Tapia, P., Rodríguez, M., Kloss, S., Cofré-Morales, C., Ormeño-Arriagada, P., & Huencho-Iturra, K. (2024). *Automated identification and analysis of Barrett's Taxonomy levels in reading comprehension assessment tasks: A GPT-based*

approach. Universidad Viña del Mar / Universidad de Playa Ancha / Universidad de Valparaíso.

Rastle, K. (2019). The place of morphology in learning to read in English. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, 116, 45–54.
<https://doi.org/10.1016/j.cortex.2018.02.008>

Ready, C. H. (2014). Tensiones y desafíos para pensar el cambio en la formación práctica de futuros profesores. *Estudios Pedagógicos*, 40(Especial), 127-143.
<https://doi.org/10.4067/s0718-07052014000200008>

Reid, G. (2016). *Dyslexia: A Practitioner's Handbook*. John Wiley & Sons.

Rodríguez, C. C., Leal, X. P., & Farías, C. S. (2021). Docentes emocionalmente inteligentes. Importancia de la Inteligencia Emocional para la aplicación de la Educación Emocional en la práctica pedagógica de aula. *Estudios Pedagógicos*, 47(1), 219-233. <https://doi.org/10.4067/s0718-07052021000100219>

Rodríguez Pech, J., & Chan Chi, G. I. (2023). Innovación educativa y formación del profesorado, una revisión estándar de literatura. *Foro educacional*, 40, 7–37.
<https://doi.org/10.29344/07180772.40.33>

Sáiz, J. (2014). Fuentes históricas y libros de texto en secundaria: Una oportunidad perdida para enseñar competencias de pensamiento histórico. *ENSAYOS. Revista de la Facultad de Educación de Albacete*, 29(1), 83-99.
<https://revista.uclm.es/index.php/ensayos/article/view/503/458>

Schank, R. C. (1975). *El papel de la memoria en el procesamiento del lenguaje, Estructura de la memoria humana*. Omega.

Schiefelbein, E. (2008). Desafíos para los investigadores en educación de América Latina. *Magis. Revista Internacional de Investigación en Educación*, 1(1), 31-44.

Solé, I. (2006). *Estrategias de lectura*.

Solé, I. (2004). Proyectos y programas de innovación en la enseñanza y el aprendizaje de la lectura y la escritura. *La práctica psicopedagógica en educación formal*, 1(1), 253–274

Subramani, M., Jaleel, I., & Mohan, S. K. (2023). Evaluating the performance of ChatGPT in medical physiology university examination of phase I MBBS. *Advances In Physiology Education*, 47(2), 270-271. <https://doi.org/10.1152/advan.00036.2023>

Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments*, 10(1). <https://doi.org/10.1186/s40561-023-00237-x>

Universidad de Chile. (2011). *Evaluación de los resultados en comprensión lectora en educación básica*. Facultad de Educación, Universidad de Chile. Recuperado de <http://www.universidaddechile.cl/estudio2011>

Anexos

1. Material de apoyo para el desarrollo del cuestionario.

Estimado/a participante:

El presente cuestionario tiene como propósito recoger su percepción sobre los textos y reactivos de comprensión lectora que se generaron usando ChatGPT, teniendo como referencia: Los objetivos de aprendizaje, los estándares de aprendizaje y la taxonomía de Barret. Para ello, necesitamos que evalúe la pertinencia, calidad, consideraciones prácticas, y aspectos éticos y de justicia de cada texto y respectivos reactivos.

Objetivos de Aprendizaje (OA):

- OA 2: Identificar y extraer información explícita en textos literarios y no literarios.
- OA 4: Inferir información a partir de los textos, relacionando elementos explícitos con conocimientos previos.
- OA 5: Interpretar y analizar el lenguaje figurado en textos literarios para comprender el mensaje del autor.
- OA 6: Evaluar críticamente textos, estableciendo juicios de valor sobre el contenido o la intención comunicativa del autor.

Estándares de Aprendizaje:

Los estándares utilizados corresponden a los orientados a la evaluación SIMCE para cuarto básico, los cuales establecen expectativas claras en habilidades como:

- Comprender información explícita e implícita.
- Relacionar elementos del texto con su contexto sociocultural.
- Emitir juicios críticos y reflexivos sobre el contenido y la estructura del texto.

Niveles de la Taxonomía de Barret:

- Comprensión literal: Reconocimiento de datos e información explícita en el texto.

- Comprensión inferencial: Relación entre el contenido del texto y el conocimiento previo del lector.
- Comprensión crítica: Evaluación del texto, emitiendo juicios basados en criterios personales o externos, considerando la intención del autor y el contexto del mensaje.

Para completar este cuestionario, se solicita evaluar los textos y reactivos según los siguientes criterios:

Pertinencia:

- Evalúa si los textos seleccionados y los reactivos diseñados están alineados con las habilidades de comprensión lectora definidas en el currículo nacional para cuarto básico. Este criterio considera si los textos y reactivos son relevantes y adecuados para el nivel educativo de los estudiantes.

Calidad:

Este criterio incluye la validez y la confiabilidad de los reactivos:

- Validez: Determina si los reactivos evalúan de manera precisa las habilidades de comprensión lectora estipuladas en los objetivos de aprendizaje.
- Confiabilidad: Analiza si los reactivos pueden ser aplicados en diferentes contextos de manera consistente, proporcionando resultados confiables.

Consideraciones prácticas:

Este criterio analiza aspectos relacionados con la implementación de los reactivos:

- Tiempo: Evalúa si los reactivos pueden ser completados en el tiempo disponible dentro de una clase típica.
- Facilidad de aplicación: Determina si los reactivos son sencillos de administrar y analizar para el docente.

- Recursos: Considera si los recursos necesarios para implementar los reactivos son accesibles en un contexto escolar promedio.

Aspectos éticos y de justicia:

Este criterio valora si los textos y reactivos cumplen con los principios de equidad y accesibilidad:

- Neutralidad del lenguaje: Evalúa si el lenguaje utilizado es inclusivo y libre de sesgos culturales, lingüísticos o socioeconómicos.
- Adaptabilidad: Considera si los reactivos pueden ajustarse para atender las necesidades de estudiantes con discapacidades o necesidades educativas especiales.
- Justicia: Garantiza que los reactivos no presenten desventajas para ningún grupo específico de estudiantes.

2. Textos seleccionados por ChatGpt y sus respectivos reactivos desarrollados por la IA.

Texto 1: La liebre y la tortuga

Jean de La Fontaine

Una liebre y una tortuga hicieron una apuesta. La tortuga dijo:

—Te apuesto que no llegas tan pronto como yo a ese árbol...

—Estás loca —contestó la liebre riendo.

—Loca o no, mantengo la apuesta.

Apostaron, y pusieron junto al árbol lo apostado. No interesa a nuestro cuento saber lo que era ni tampoco quién fue el juez de la contienda.

Nuestra liebre no tenía que dar más que cuatro saltos. Ella los da muy contenta, sus patas apenas se ven devorando el campo y la pradera.

Tenía, pues, tiempo de sobra para mordisquear la hierba, para dormir y para olfatear el viento.

Dejó a la tortuga partir con su pasito calmo. Esta partió lentamente.

La liebre, en tanto, desdeñando una fácil victoria, tuvo en poco a su contrincante, y juzgó más decoroso no emprender la carrera hasta última hora. Se quedó tranquila sobre la fresca hierba, y se entretuvo atenta a cualquier cosa, menos a la apuesta. Cuando vio que la tortuga llegaba ya a la meta, partió como un rayo; pero fue inútil. La primera que llegó a la meta fue su rival.

—¿Qué te parece? —le dijo la tortuga—. ¿Tenía o no tenía razón? ¿De qué te sirve tu agilidad siendo tan presumida? ¡Vencida por mí! ¿Qué te pasaría, si llevaras, como yo, la casa a cuestas?

Reactivos de comprensión.

1. ¿Qué animales compiten en la carrera y quién resulta ganador?
2. ¿Qué hace la liebre después de adelantar a la tortuga?
3. ¿Por qué la tortuga no se rinde a pesar de ir más lenta que la liebre?
4. ¿Qué enseñanzas podemos sacar sobre el esfuerzo al analizar el comportamiento de la

tortuga?

5. ¿Crees que el comportamiento de la liebre fue justo? ¿Por qué?

Texto 2: El vendedor de sueños

María Elena Walsh

Vendo sueños con gusto a
caramelito,

países raros, lentas
maravillas, ángeles que dan
cine por el cielo, y relámpagos
para pesadillas.

Sueños como trapitos de
colores, imágenes y muchas
otras cosas. Algunos tienen
pájaros y flores. Otros, infierno y
brujas espantosas.

Sueños y sueños para todo
gusto:

cajas de azufre, paquetitos
rojos. Lágrimas o canción, amor
o susto para los niños que
cierran los ojos.

Llevo en mi cesta el mágico
tesoro.

¡A ver quién me lo compra,
quién me llama!

Dejen afuera su moneda de
oro, y mírenme pasar desde
la cama.

Reactivos de comprensión.

1. ¿Qué objetos o elementos menciona el poeta que el vendedor de sueños ofrece?
2. ¿Cómo describe el poema al vendedor de sueños? ¿Qué hace él?
3. ¿Qué crees que representan los sueños en este poema?
4. ¿Por qué el poeta dice que los sueños tienen valor, aunque no sean reales?
5. ¿Crees que es importante "comprar sueños" en la vida real? ¿Por qué?

Texto 3: El pasaje de la Oca

Elsa Bornemann

El pasaje de la Oca era una callecita muy angosta... Tan angosta que a las personas que allí vivían les bastaba estirar las manos a través de las ventanas para estrechar las de los vecinos de enfrente. Todos eran felices allí y yo no tendría nada que contarles si una madrugada no hubiera llegado al Pasaje de la Oca el señor Álvaro Rueda. Este señor estacionó su automóvil justo a la entrada del pasaje y tocó insistentemente la poderosa bocina hasta despertar a los habitantes de la callecita. En cinco minutos ya estaban todos alrededor del auto, entre dormidos y asustados, preguntándole qué sucedía.

Álvaro Rueda, mostrándoles un plano, les anunció la terrible noticia:

—Señores vecinos, yo soy el dueño de este terreno. Lamento comunicarles que la semana próxima desaparecerá el Pasaje de la Oca. Haré demoler todas las casas, puesto que aquí construiré un gran edificio para archivar mi valiosa colección de estampillas... Múdense cuanto antes— y, despidiéndose con varios bocinazos, puso en marcha su vehículo y se perdió en la avenida. Por un largo rato, los vecinos del Pasaje de la Oca no hablaron, no lloraron ni se movieron: tanta era su sorpresa. Parecían fantasmas dibujados por la luna, con sus camiones agitándose con el viento del amanecer. Más tarde, sentándose en los cordones, estudiaron diferentes modos de salvar el querido pasaje:

- Desobedecer al señor Rueda y quedarse allí por la fuerza. Pero esta solución era peligrosa: ¿Y si Álvaro Rueda, furioso, ordenaba lanzar las máquinas topadoras sobre el pasaje, sin importarle nada? No. En ese caso, lo perderían sin remedio...
- El Pasaje de la Oca podría ser enrollado como una alfombra y trasladado a otra parte; solución que fue descartada:
 - ¡No! ¡Imposible! ¡Se quebrarían todas las copas! ¡Se harían añicos las jarras y los floreros de vidrio! ¿Cómo salvarían los espejos?
- Podrían contratar a un hechicero de la India para que colocara el pasaje sobre una

alfombra voladora y lo llevara, por el aire, a otra región. Pero la India estaba lejos de allí... y el viaje por avión costaba demasiado dinero...

Ya estaban por darse por vencidos, resignándose a perder su querida callecita, cuando el anciano don Martín tuvo una idea sensacional:

—¡Viva! ¡Encontré la solución! Escuchen: nos dividiremos en dos grupos y cada uno tomará el pasaje por un extremo. Los de adelante tirarán de la calle con todas sus fuerzas y los de atrás empujarán con vigor. De ese modo, podremos despegarla y llevarla, arrastrando, hasta encontrar un terreno libre donde colocarla otra vez. ¡El Pasaje de la Oca no será destruido!

—¡Viva don Martín! —gritaron todos los vecinos, contentísimos. Y esperaron la noche para realizar su extraordinario plan.

Fue así como, cuando toda la ciudad dormía, los habitantes del Pasaje de la Oca lo tomaron de las puntas y empezaron la mudanza. Despegar el pasaje fue lo que más trabajo les costó, porque arrastrarlo no resultó dificultoso. El pasaje se dejaba llevar como deslizándose sobre una pista encerada. Pronto encontraron la avenida, suficientemente ancha como para permitir el paso de la callecita... Y allá fueron todos, hombres, mujeres y niños, llevándose el pintoresco pasaje a cuestas, como un maravilloso teatrillo ambulante, con sus casitas blancas y humildes bamboleándose durante la marcha, con sus farolitos pestañando luces amarillentas, con sus sábanas bailando en las sogas de las terrazas bajo un pueblito de estrellas echado boca abajo.

La mañana siguiente abrió sus telones y vio al Pasaje de la Oca instalado en el campo. Allí, sobre el chato verde, lo colocaron felices. Esa noche celebraron una gran fiesta y los fuegos artificiales estrellaron aún más la noche campesina. A la mañana siguiente, cuando el señor Álvaro Rueda llegó, seguido por una cuadrilla de obreros dispuestos a demoler el pasaje, encontró el terreno completamente vacío.

—¡El callejón desapareció! —alcanzó a gritar antes de caer desmayado. Y nunca supo que la generosidad del campo había recibido al pasaje, callecita fundadora del que, con el correr del tiempo, llegó a ser el famoso PUEBLO DE LA OCA.

Reactivos

1. ¿Dónde transcurre la historia y quiénes son los personajes principales?
2. ¿Qué sucede cuando los personajes se llevan el pasaje de la Oca?
3. ¿Por qué crees que los personajes deciden llevarse el pasaje a pesar del peligro?
4. ¿Cómo influye el entorno en las decisiones de los personajes?
5. ¿Crees que las decisiones de los personajes fueron correctas? ¿Qué habrías hecho tú en su lugar?

3. Carta a expertos para validación de preguntas del cuestionario

Valparaíso, 4 de noviembre 2024

Validación de Instrumento

Mi nombre es Patricio Tapia Leyton, soy profesor de educación general básica y estudiante del magíster en Evaluación Educacional de la Universidad de Playa Ancha. En el marco de mi tesis, vengo a solicitar a usted su ayuda para validar preguntas de una escala likert que se utilizarán para recabar información sobre las percepciones que hará un grupo de profesores sobre un conjunto de textos que seleccionó una Inteligencia Artificial (IA), ChatGPT, de un texto escolar para cuarto año básico¹ y la respectiva serie de reactivos que dicha IA generó para cada uno de los textos que seleccionó.

1.-Objetivo de la validación.

La validación busca obtener retroalimentación sobre la claridad, relevancia y pertinencia de las preguntas de la escala Likert, que se encuentran agrupadas en cuatro categorías y que van en atención del conjunto de textos seleccionados y los reactivos generados mediante ChatGPT.

2.-Estrategia Metodológica para la selección de textos y generación de reactivos

Se crearon *prompts* para instruir a la IA en la selección de los textos del texto escolar y la creación de preguntas de comprensión lectora utilizando para ello los siguientes criterios de referencia:

- i. Taxonomía de Barret (comprensión literal, comprensión inferencial y crítica)
- ii. Objetivos de Aprendizaje del programa de estudio de cuarto básico², enfocados en la comprensión lectora (OA2 - OA4 - OA5 - OA6)
- iii. Estándares de aprendizaje de cuarto básico³.

Como primera instrucción se entregó un *prompts* para contextualizar y entregar información sobre el tipo de textos y las características de los respectivos reactivos: la descripción de tres niveles de comprensión lectora según la taxonomía de Barret, los objetivos de aprendizaje relacionados a la comprensión lectora y los estándares de aprendizaje ministeriales para cuarto año básico.

Primera indicación a ChatGPT

A continuación, se adjuntan los documentos sobre el programa ministerial, estándares de aprendizaje y el libro del estudiante del MINEDUC, todo esto para enriquecer la memoria de la IA.

Posterior a la entrega de toda esta información, se le pidió lo siguiente:

¹ Briceño V., C., González M., V., Jiménez A., C., & Rolando H., C. (2023). *Lenguaje y comunicación. Texto del estudiante, 4º básico* (Edición especial para el Ministerio de Educación). Universidad San Sebastián.

² Ministerio de Educación de Chile. (2018). *Lenguaje y comunicación: Programa de estudio para cuarto año básico* (2.ª ed.). Unidad de Currículum y Evaluación.

³ Ministerio de Educación de Chile. (2019). *Estándares de aprendizaje: Lectura 4º básico* (Actualización 2019). Santiago de Chile.

"Selecciona 4 textos desde el libro del estudiante, los cuales sean los más idóneos para la creación de diversos reactivos centrados en los Objetivos de aprendizaje de cuarto, la taxonomía de Barret y los estándares de aprendizaje de este nivel. Estos textos deben ser: fabula, poema, texto informativo y cuento"

Y, por último, después de la selección de textos realizada se entregó el siguiente *prompts*:

"Necesito que desarrolles cinco reactivos evaluativos para cada uno de esos textos que se centren en los tres niveles de la taxonomía de Barret (literal, inferencial, crítica). Relaciona cada reactivo con los Objetivos de Aprendizaje (OA2, OA4, OA5 Y OA6) correspondientes del currículum nacional de Lenguaje para cuarto básico, y con los Estándares de Aprendizaje del SIMCE para este nivel. Además, justifica cada reactivo en función del nivel de la taxonomía de Barret, el OA, y el estándar correspondiente"

Categorías

Las preguntas que responderán los docentes se agrupan en cuatro categorías, las cuales incluyen un conjunto de características que debe cumplir un instrumento evaluativo con sus respectivos reactivos Para realizar una evaluación

Coherencia:

Es la alineación entre el objetivo de evaluación y el tipo de instrumento utilizado. Un instrumento coherente obtiene información relevante sobre el progreso de los estudiantes en las habilidades evaluadas. La coherencia asegura que los reactivos correspondan a los propósitos de la evaluación (diagnóstico, logro, etc.) y se ajusten al contenido del currículum nacional y a las características de la población objetivo.

Calidad:

Es la característica de los reactivos que asegura dos aspectos clave: (a) validez, es decir, que el instrumento evalúe los contenidos y habilidades deseadas; y (b) confiabilidad, o que los resultados sean consistentes bajo condiciones similares. La calidad garantiza que el instrumento mide lo que se pretende y ofrece resultados confiables en contextos comparables.

Consideraciones Prácticas:

Se refiere a la viabilidad de la implementación en términos de costo, tiempo y facilidad de administración. Un instrumento práctico necesita recursos accesibles y un tiempo adecuado para el contexto escolar. Además, el proceso de puntuación y análisis debe ser eficiente y aplicable a nivel grupal dentro del aula.

Ética y Justicia:

Se refiere a asegurar un proceso de evaluación justo para todos los estudiantes, considerando sus contextos culturales y necesidades específicas. Los reactivos deben diseñarse sin sesgos y adaptarse para estudiantes con discapacidades u otras necesidades especiales. La ética y justicia garantizan que cada estudiante tenga igualdad de oportunidades para demostrar su comprensión, sin limitaciones por factores externos como el idioma, el contexto cultural o las condiciones físicas.

4. Cuestionario.

Estimado/a participante:

El presente cuestionario tiene como propósito recoger su percepción sobre los reactivos de comprensión lectora que se generaron usando ChatGPT, teniendo como referencia: Los objetivos de aprendizaje, los estándares de aprendizaje y la taxonomía de Barret. Para ello, necesitamos que evalúe la pertinencia, calidad, consideraciones prácticas, y aspectos éticos y de justicia de cada texto y respectivos reactivos.

Lea cuidadosamente los textos proporcionados y responda cada ítem según su percepción. Utilice la escala para calificar cada texto. Puede agregar comentarios adicionales en la sección correspondiente.

Se solicita que responda cada ítem marcando en la opción que mejor represente su opinión en una escala del 1 al 5:

- 1 = Muy inadecuado
- 2 = Inadecuado
- 3 = Indiferente
- 4 = Adecuado
- 5 = Muy adecuado

- Descargue aquí los documentos necesarios:

1. OA, Estándares, Taxonomía y criterios de evaluación:
https://drive.google.com/file/d/1ZruHBnEQbD_wbtaigpc0sw2pofvNXqd_/view?usp=drive_link
2. Textos y sus reactivos (Material seleccionado y creado por Chatgpt):
https://drive.google.com/file/d/1eci4erZkhRKiPPmVF9v-Aa2ZspV5Rg16/view?usp=drive_link

Criterio Coherencia.

Pregunta 1: ¿Considera que los textos seleccionados están alineados con las habilidades de comprensión lectora definidas para los estudiantes de cuarto año básico?

Pregunta 2: ¿Considera que los reactivos creados para cada texto permiten obtener información relevante sobre el nivel de comprensión lectora de los estudiantes de cuarto año básico según los niveles de la taxonomía de Barret?

Pregunta 3: ¿Considera que los reactivos propuestos por ChatGPT cubren adecuadamente los contenidos y habilidades establecidos en el currículo para el nivel de cuarto año básico?

Pregunta 4: ¿Considera que los reactivos creados por ChatGPT son apropiados para los estudiantes de cuarto básico, teniendo en cuenta las características y necesidades educativas según tu contexto o experiencia?

Pregunta 5: ¿Cree que los reactivos propuestos por ChatGPT reflejan adecuadamente los niveles de desempeño esperados en comprensión lectora para los estudiantes de cuarto básico según los objetivos de aprendizaje?

Criterio calidad

Pregunta 6: ¿Considera que la redacción de los reactivos es adecuada para evaluar de manera válida y confiable la comprensión lectora de los estudiantes de cuarto año básico?

Pregunta 7: ¿Cree que la aplicación de este conjunto de reactivos en su contexto laboral proporcionaría resultados confiables y válidos sobre el nivel de comprensión lectora de los estudiantes?

Pregunta 8: ¿Considera que el conjunto de reactivos es consistente y válido en la medición de las habilidades de comprensión lectora definidas en el currículo nacional?

Pregunta 9: ¿Considera que los textos seleccionados son válidos y apropiados para evaluar de manera confiable las habilidades de comprensión lectora en estudiantes de cuarto básico?

Criterio Consideraciones prácticas.

Pregunta 10: ¿Considera que el conjunto de textos y sus reactivos es adecuado para ser aplicado en un bloque de 90 minutos, teniendo en cuenta el tiempo necesario para su administración y análisis?

Pregunta 11: ¿Considera que el análisis de los resultados de cada reactivo requiere un esfuerzo razonable en términos de tiempo y conocimientos por parte del docente?

Pregunta 12: ¿Considera que el conjunto de reactivos se puede administrar de manera eficiente en un grupo de 30 estudiantes, como el promedio en Chile, en términos de tiempo y recursos disponibles?

Criterio Ética y justicia.

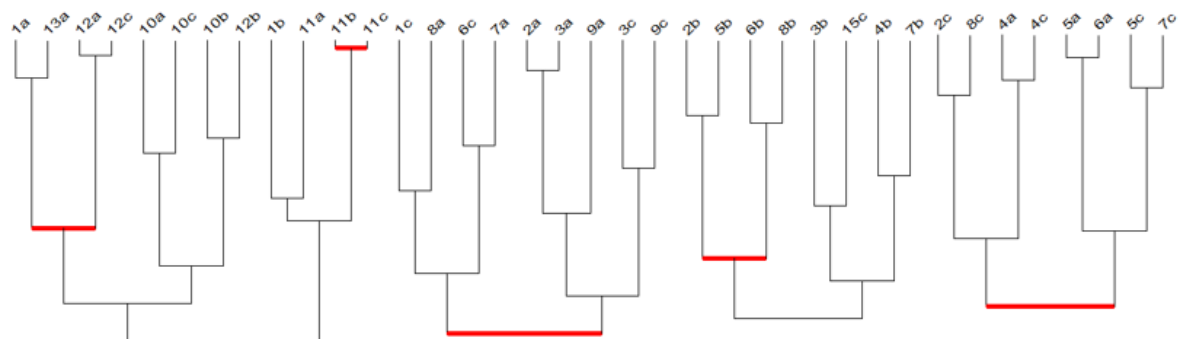
Pregunta 13: ¿Considera que el lenguaje utilizado en los textos y reactivos es neutral y accesible para todos los estudiantes, incluyendo aquellos de diferentes contextos culturales y niveles socioeconómicos?

Pregunta 14: ¿Considera que los reactivos evitan sesgos que puedan afectar la evaluación de estudiantes de diferentes orígenes culturales o lingüísticos, garantizando un trato justo para todos?

Pregunta 15: ¿Considera que los reactivos pueden ser diversificados o adaptados para atender las diversas necesidades educativas de los estudiantes en su contexto laboral, incluyendo estudiantes con discapacidades o necesidades especiales?

5. Resultados de la aplicación de cuestionario a docentes.

a) Árbol de similaridad



Clases Significativas

Clasificación del nivel : 1 : (11b 11c) similaridad : 0.744507

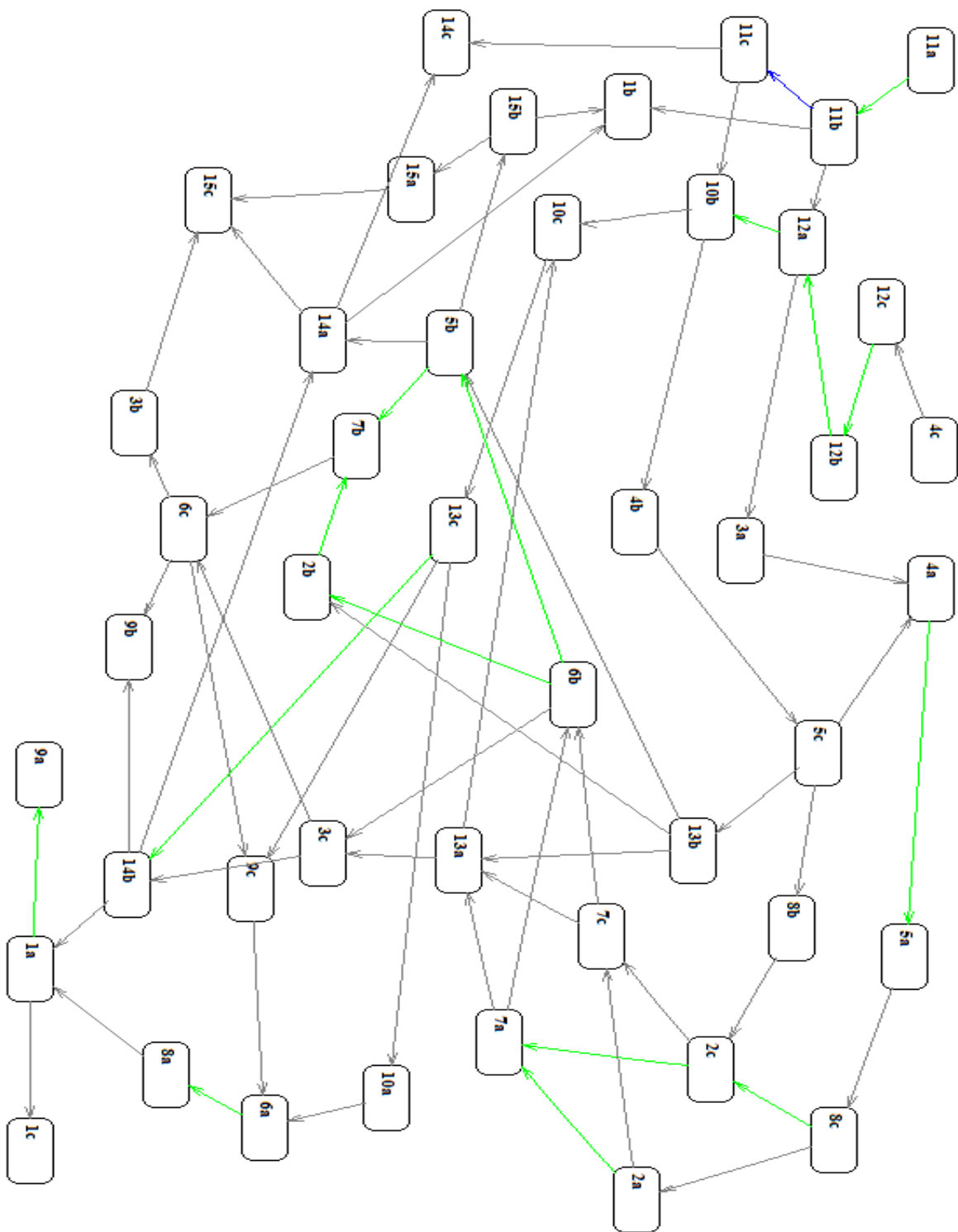
Clasificación del nivel : 25 : ((1a 13a) (12a 12c)) similaridad : 0.159729

Clasificación del nivel : 29 : ((2b 5b) (6b 8b)) similaridad : 0.131404

Clasificación del nivel : 36 : (((2c 8c) (4a 4c)) ((5a 6a) (5c 7c))) similaridad : 0.000519032

Clasificación del nivel : 39 : (((1c 8a) (6c 7a)) (((2a 3a) 9a) (3c 9c))) similaridad : 5.25414e-05

b) Grafo Implicativo



c) Árbol Cohesitivo

