

# Desarrollo gráfico expresivo en Educación Infantil: análisis comparativo entre evaluadores humanos y ChatGPT-4o

## Expressive Graphic Development in Early Childhood Education: A Comparative Analysis of Human Evaluators and ChatGPT-4o

Jesica León Paredes<sup>1</sup>, Jesús Antonio Gómez Ochoa de Alda<sup>2</sup>, Cristina Valares Masa<sup>3</sup>, José María Marcos Merino<sup>4</sup>, María Rocío Esteban Gallego<sup>5</sup>

<sup>1</sup> Universidad de Extremadura. Cáceres [jleonpar@alumnos.unex.es](mailto:jleonpar@alumnos.unex.es)

<sup>2</sup> Universidad de Extremadura. Cáceres [ochoadealda@unex.es](mailto:ochoadealda@unex.es)

<sup>3</sup> Universidad de Extremadura. Cáceres [cvalmas@unex.es](mailto:cvalmas@unex.es)

<sup>4</sup> Universidad de Extremadura. Badajoz [jmmarcos@unex.es](mailto:jmmarcos@unex.es)

<sup>5</sup> Universidad de Salamanca. Salamanca [rocioesteban@usal.es](mailto:rocioesteban@usal.es)

Recibido: 2/10/2025

Aceptado: 10/4/2026

Copyright ©

Facultad de CC. de la Educación y Deporte.  
Universidad de Vigo



Dirección de contacto:

Cristina Valares Masa

Dpto. Didáctica de Ciencias

Experimentales y Matemáticas

Facultad de Formación del Profesorado

Universidad de Extremadura

Avda. de la Universidad, s/n

10003 Cáceres

### Resumen

El dibujo infantil es un medio de comunicación no verbal que integra dimensiones motrices, cognitivas, emocionales y sociales, y su análisis facilita al docente el seguimiento del desarrollo integral del alumnado de Educación Infantil. Este estudio diseña y valida un instrumento, basado en la literatura especializada, para evaluar cuatro dimensiones (motricidad fina, desarrollo cognitivo y de representación, indicadores emocionales y dimensión social/autoimagen). Se analizaron 22 dibujos de niños de 3-4 años, valorados por cinco expertos humanos y cuatro instancias del modelo generativo ChatGPT-4o. Las calificaciones humanas mostraron alta consistencia interna ( $\alpha > 0,83$ ), mientras que las de ChatGPT-4o evidenciaron baja fiabilidad ( $\alpha < 0,44$ ) y escasa coherencia interevaluador. El análisis factorial confirma la validez de la escala como indicador del desarrollo gráfico-expresivo. Se concluye que, aunque ChatGPT-4o aporta ventajas operativas, su sensibilidad interpretativa no permite sustituir el juicio experto; se recomienda su empleo únicamente como apoyo bajo supervisión profesional.

### Palabras clave

Educación Infantil, Desarrollo Gráfico, Dibujos Infantiles, Inteligencia Artificial, GPT-4o

### Abstract

Children's drawings constitute a non-verbal means of communication that integrates motor, cognitive, emotional, and social dimensions. Their systematic assessment enables early-childhood teachers to monitor pupils' holistic development. This study designs and validates an instrument, grounded in scientific literature, to evaluate four

dimensions: fine-motor skills, cognitive and representational development, emotional indicators, and social/self-image. Twenty-two drawings by 3- to 4-year-olds were assessed by five human experts and four instances of the generative language model ChatGPT-4o. Human ratings exhibited high internal consistency ( $\alpha > 0,83$ ); ChatGPT-4o's ratings showed low reliability ( $\alpha < 0,44$ ) and poor inter-instance agreement. Factor analysis supports the validity of the scale as an indicator of expressive-graphic development. Although ChatGPT-4o offers operational advantages, its limited interpretive sensitivity precludes replacing expert judgement; its use should therefore remain complementary under professional supervision.

### **Key Words**

Early Childhood Education, Graphic Development, Children's Drawings, Artificial Intelligence, GPT-4o

---

## **1. INTRODUCCIÓN**

El dibujo constituye, antes de la plena adquisición del lenguaje oral, un canal privilegiado de expresión durante la infancia. A través de trazos, colores y composiciones, los niños comunican emociones, ideas y vivencias que aún no pueden verbalizar, configurando un lenguaje visual con significado propio (Molina-Jiménez, 2015). Esta capacidad expresiva convierte al dibujo en una herramienta esencial durante el desarrollo temprano, facilitando la exteriorización del pensamiento y evolucionando hacia formas más comprensibles para el adulto. Lejos de ser meros garabatos, estas producciones articulan dimensiones motrices, cognitivas, emocionales y sociales (Cox, 2005; Farokhi y Hashemi, 2011), y proporcionan al docente indicios valiosos sobre el desarrollo integral del alumnado.

Interpretar adecuadamente los dibujos infantiles en el ámbito educativo constituye una tarea tan relevante como compleja. No se trata de identificar “dibujos mejores” en términos absolutos, sino de comprender que cada producción refleja la realidad interna del niño y la manera en que organiza perceptivamente su entorno. Este análisis requiere una mirada experta, que combine sensibilidad estética, conocimiento del desarrollo gráfico-expresivo y comprensión del contexto cultural de cada niño (Wright, 2007; Rinaldi, 2006). Dado que no existen códigos universales de interpretación, se precisa una observación y una escucha atentas, sustentadas en una sólida formación específica que permita al docente de Educación Infantil acompañar de forma adecuada el proceso expresivo del alumnado, reconociendo la singularidad de cada obra y el contexto en que se genera (Wright, 2007; Rinaldi, 2006; Lowenfeld y Brittain, 1987).

### **1.1. Valor pedagógico del dibujo en Educación Infantil**

Diversos modelos teóricos clásicos describen el desarrollo gráfico infantil –entre los 18 meses y los 6 años– como un proceso evolutivo que transita desde el garabateo inicial hasta formas progresivamente más organizadas y representativas, como el simbolismo y el realismo visual. Este proceso se evidencia en una mayor coherencia interna, organización espacial y riqueza expresiva en los dibujos (Lowenfeld y Brittain, 1987; Mestre, 2010; Quiroga, 2007). A su vez, estos productos gráficos reflejan el grado de maduración motriz fina, el desarrollo de nociones espaciales y la capacidad de abstracción. Elementos como la elección cromática, la presión del trazo o la aparición de

figuras significativas pueden ofrecer indicios sobre el estado emocional del niño y sus vínculos sociales (Moreno, 2022).

Desde una perspectiva constructivista del aprendizaje, el dibujo cumple una función diagnóstica fundamental: permite acceder a las ideas previas del alumnado, identificar concepciones intuitivas o espontáneas, y detectar posibles errores en la comprensión de determinados contenidos curriculares. Esta información es clave para que el docente planifique su intervención didáctica desde el desarrollo real de cada niño, favoreciendo así una construcción progresiva y significativa del conocimiento (Quiroga, 2007; Vygotsky, 1978). A esta función diagnóstica se suma la dimensión expresiva, que convierte al dibujo en un canal simbólico difícilmente sustituible. Como herramienta pedagógica en Educación Infantil, el dibujo permite no solo estimular el desarrollo gráfico-expresivo, sino también favorecer el crecimiento cognitivo, emocional, físico y social del alumnado. Dibujar es una forma de proyectar el mundo interno, lo que facilita una comprensión más profunda por parte del docente y una intervención educativa más ajustada y efectiva (Cherney et al., 2006; Zlateva, 2019).

Así, el uso del dibujo como estrategia educativa, se revela como una vía eficaz para explorar las concepciones infantiles en torno a fenómenos científicos y guiar la enseñanza a partir de ellas. En consecuencia, el análisis de los dibujos no solo debe circunscribirse a una interpretación psicológica, sino que también debe considerarse una herramienta metodológica que potencia la comprensión del pensamiento infantil y favorece una enseñanza más ajustada, reflexiva y contextualizada.

## 1.2. Criterios para el análisis gráfico-expresivo

El análisis del dibujo permite observar la organización, el uso del espacio, la percepción, las emociones y rasgos de personalidad. Características como la presión del trazo, el tamaño de las figuras, la elección de colores y la inclusión de detalles anatómicos ofrecen indicios interpretativos (Cherney et al., 2006; Farokhi y Hashemi, 2011).

Diversos estudios coinciden en que factores como los materiales, el tamaño del papel, de las figuras y del trazo revelan emociones, personalidad y necesidades. Por ejemplo, el uso de lápices gruesos y colores vivos indican seguridad; los lápices delgados y tonos apagados, inseguridad (Zlateva, 2019; Moreno, 2022). Del mismo modo, figuras de gran tamaño y que ocupan mucho espacio reflejan extroversión, impulsividad y seguridad, mientras que formatos pequeños y centrados pueden implicar timidez o inseguridad (Lowenfeld y Brittain, 1987), aunque estas interpretaciones deben considerarse orientativas y contextualizadas dentro del desarrollo evolutivo del niño. En cuanto al trazo, una presión fuerte puede interpretarse como manifestación de tensión muscular o ansiedad, mientras que los trazos débiles o discontinuos podrían reflejar vergüenza, inhibición o baja energía emocional. El sombreado excesivo puede ser signo de ansiedad; si es localizado, la ansiedad podría relacionarse con esa zona (Farokhi y Hashemi, 2011; Moreno, 2022; Zlateva, 2019). Estas variables gráficas, lejos de ser meros aspectos técnicos, pueden construir indicadores significativos del mundo interno del niño y deben ser interpretadas con sensibilidad y conocimiento especializado.

De igual modo, algunos detalles anatómicos también aportan información relevante. Una cabeza desproporcionada indica egocentrismo y si es equilibrada, desarrollo emocional normal. Rasgos como una boca grande o dientes marcados reflejan agresividad; ojos grandes, vigilancia; pequeños, dependencia. Una nariz proporcional es

señal de desarrollo adecuado; fosas nasales grandes, de agresividad. Cuellos largos indican dificultad para alcanzar metas; su ausencia, inseguridad. Manos extendidas expresan deseo de conexión; grandes, impulsividad; pequeñas, timidez (Farokhi y Hashemi, 2011).

La interpretación de los dibujos infantiles puede ayudar a detectar baja autoestima, dificultades emocionales e incluso maltrato (Moreno, 2022). Dicha interpretación requiere de experiencia, observación minuciosa, análisis riguroso y consulta de bibliografía. De este modo, el análisis de los dibujos debe realizarse con precaución y siempre por especialistas capacitados. Pese a la abundancia de criterios dispersos, actualmente no existen rúbricas cualitativas estandarizadas que permitan codificar de manera sistemática las producciones infantiles (Campbell y Cunningham, 2025), lo que puede incrementar la subjetividad en las interpretaciones. De ahí la necesidad de obtener instrumentos validados que garanticen fiabilidad y reduzcan el sesgo interpretativo.

### **1.3. La Inteligencia Artificial y los grandes modelos de lenguaje**

Un modelo de lenguaje grande (Large Language Model, LLM) es un sistema de Inteligencia Artificial (IA) que ha sido entrenado para entender y generar texto, y aprender qué palabras suceden a otras basándose en patrones extraídos de grandes cantidades de texto. ChatGPT (Generative Pre-Trained Transformer), de OpenAI, es uno de estos modelos que ha sido diseñado para mantener un diálogo de forma natural (asistente de diálogo o chatbot) y utiliza un tipo de arquitectura de IA denominada Transformer (Bubeck et al., 2023). Este y otros asistentes de diálogo (como Deepseek, Gemini o Copilot) están siendo aplicados con éxito en entornos educativos, ofreciendo apoyo personalizado y adaptado a las necesidades individuales (Ayuso y Gutiérrez, 2022). La IA puede analizar grandes volúmenes de datos simultáneamente, operando sin descanso y mejorando la eficiencia en diversos ámbitos, incluidos salud, educación y trabajo (Rouhiainen, 2018).

En el ámbito educativo, su aplicación está respaldada por organismos como la UNESCO y está vinculada a los Objetivos de Desarrollo Sostenible (ODS), ya que favorece el acceso al conocimiento, la retroalimentación precisa y la personalización del aprendizaje (Granda et al., 2024). La IA ofrece beneficios tanto académicos como emocionales en los estudiantes y docentes, tales como explorar en un mundo amplio de recursos educativos o asistencia digital en el ámbito afectivo (Ayuso y Gutiérrez, 2022). Algunos estudios recientes han demostrado que modelos como GPT pueden ofrecer resultados consistentes y válidos en tareas de análisis psicológico multilingüe, lo que evidencia su potencial para apoyar evaluaciones complejas que requieren interpretación de contenidos subjetivos (Rathje et al., 2024). De este modo, se ha propuesto que los LLM puedan llevar a cabo evaluaciones sistemáticas y objetivas, reduciendo la subjetividad inherente al juicio humano (Aparicio, 2023). Sin embargo, persisten dudas sobre su capacidad para igualar la sensibilidad interpretativa del juicio humano (García Peña et al., 2020), especialmente en contextos como la Educación Infantil, donde la comprensión emocional y contextual resulta crucial.

Los desarrollos más recientes de ChatGPT, como el modelo GPT-4o, comprenden y generan lenguaje, audio e imágenes en un único sistema. GPT-4o ya ha sido usado para reconocer imágenes de medicamentos y dar consejo al paciente sin errores significativos (Bazzari y Bazzari, 2024). Esto sugiere que la IA, concretamente GPT-4o, podría ser una

herramienta útil para automatizar, sistematizar y reducir la subjetividad de las evaluaciones de dibujos infantiles (Shah et al., 2025), ayudando al docente en el análisis del nivel de desarrollo del alumno.

#### **1.4. Objetivos e hipótesis**

Este estudio propone diseñar y validar un instrumento de análisis del dibujo infantil, basado en la literatura especializada, que permita evaluar de forma sistemática cuatro dimensiones fundamentales en la etapa de Educación Infantil. A partir de esta rúbrica, se contrasta la consistencia entre las valoraciones de expertos humanos y las generadas por GPT-4o, concretamente indaga: (a) en qué medida coinciden las puntuaciones emitidas por el modelo con las de los evaluadores humanos, (b) qué diferencias se observan en términos de consistencia interna y fiabilidad interevaluador, y (c) si GPT-4o puede considerarse un recurso pedagógico complementario en el aula de Educación Infantil. De acuerdo con estos propósitos, se formulan tres hipótesis: H1, la rúbrica diseñada ofrecerá resultados consistentes y fiables entre los evaluadores humanos; H2, GPT-4o mostrará un grado de concordancia moderado respecto a dichas valoraciones; y H3, existirán diferencias significativas en consistencia interna y sensibilidad interpretativa entre las puntuaciones humanas y las generadas por la inteligencia artificial.

## **2. METODOLOGÍA**

### **2.1 Diseño**

La presente investigación adopta una metodología de enfoque mixto, combinando técnicas cualitativas y cuantitativas para analizar la interpretación de dibujos infantiles desde una perspectiva subjetiva, considerando tanto la evaluación humana como la realizada por ChatGPT-4o. El estudio adopta un enfoque descriptivo y exploratorio, con el objetivo de identificar patrones, tendencias y diferencias entre ambas formas de evaluación, sin intervenir en la elaboración ni modificar el contenido de los dibujos.

### **2.2 Participantes y procedimiento**

En el estudio participaron un total de 22 alumnos del primer curso del segundo ciclo de Educación Infantil (3-4 años), en el marco de una actividad de divulgación científica del Programa Ciencia Circular de la Universidad de Extremadura, que incluye charlas y talleres para acercar la ciencia a contextos escolares. Antes de comenzar la actividad práctica del taller, se exploraron las ideas previas de los niños mediante preguntas como: «¿Sabes qué es un microorganismo? ¿Son seres vivos? Si decimos “virus” o “bacterias”, ¿qué te viene a la mente? ¿Crees que estos seres son buenos o malos?». A continuación, se les pidió que dibujaran libremente un microorganismo –virus o bacteria– según lo imaginaban; para ello recibieron una hoja A4, un lápiz y ceras de distintos colores.

La muestra, constituida por los 22 dibujos obtenidos, es de carácter intencional y responde a la disponibilidad de las producciones gráficas. Se contó con la autorización de la dirección del centro y el consentimiento informado de los padres o tutores legales. Al involucrar a menores, el procedimiento siguió los principios éticos de la Declaración de

Helsinki (Asociación Médica Mundial, 2024), garantizando voluntariedad, anonimato y confidencialidad, y fue aprobado por la Comisión de Bioética y Bioseguridad de la Universidad de Extremadura (Registro n.º 142/2025).

### 2.3. Instrumento de evaluación: diseño de la rúbrica

El instrumento central del estudio es una rúbrica elaborada específicamente para valorar el desarrollo infantil a través del dibujo. Su diseño se basó en la revisión de la literatura sobre psicología evolutiva y educación artística (Farokhi y Hashemi, 2011; Lowenfeld y Brittain, 1987; Malchiodi, 2012), así como en investigaciones que analizan la representación de microorganismos en edades tempranas (Ballesteros et al., 2018; Byrne, 2011; Domínguez et al., 2018; Faccio et al., 2013). Durante su diseño, se tomaron decisiones fundamentadas sobre la inclusión y exclusión de determinadas dimensiones evaluativas, adaptando el instrumento a los objetivos del estudio. Tras sucesivas revisiones y ajustes a los objetivos del estudio, la versión definitiva (Anexo I) quedó estructurada en cuatro dimensiones, cada una con tres o cuatro indicadores:

1. Desarrollo Motriz Fino (DMF). Evalúa habilidades motoras vinculadas al trazo gráfico.

- Control del trazo: Continuidad y dirección del trazo.
- Presión / continuidad del trazo: Uniformidad en la presión y fluidez.
- Formas básicas: Presencia de figuras geométricas simples.
- Uso del espacio: Distribución equilibrada del dibujo en la hoja.

2. Desarrollo Cognitivo y de Representación (DCR). Evalúa la capacidad de simbolizar, organizar y narrar.

- Detalle y complejidad: Cantidad y precisión de los elementos.
- Relaciones espaciales: Ubicación coherente de los elementos entre sí.
- Narrativa visual: Existencia de una historia o acción representada.

3. Desarrollo de Indicadores Emocionales (DIE). Evalúa la expresión emocional a través del color, la figura humana y su disposición.

- Uso del color: Variedad y asignación significativa de colores.
- Intensidad cromática: Saturación y sombreado expresivo.
- Figura humana: Proporción-posición e integridad anatómica (valorados como indicadores separados).

4. Dimensión Social / Autoimagen (DS). Evalúa representación de vínculos y contexto social.

- Inclusión de figuras significativas: Presencia de otras personas.
- Interacción explícita: Contacto visual, acciones compartidas.
- Elementos del fondo: Detalles ambientales o decorativos.

Cada indicador se valora en tres niveles de logro: 1 (Necesita atención); 2 (Adecuado, en el rango típico); y 3 (Avanzado). La puntuación de cada dimensión es la media aritmética de sus indicadores.

## 2.4. Evaluación de los dibujos

Los 22 dibujos se valoraron de manera paralela por dos fuentes independientes: cinco expertos humanos y el modelo GPT-4o. En cuanto a la evaluación humana (E), los profesores (E1 a E5), todos ellos docentes del Grado en Educación Infantil de las universidades de Extremadura y Salamanca, recibieron en formato PDF las veintidós imágenes escaneadas y numeradas; disponían únicamente de la rúbrica como criterio orientador y registraron sus puntuaciones en una hoja de cálculo idéntica, lo que garantizó la homogeneidad, comparabilidad de los datos y permitió el análisis estadístico.

Para la evaluación automática se crearon instancias personalizadas de GPT-4o que incluían la rúbrica, la plantilla Excel y una secuencia de instrucciones (*prompts*) que guiaban el análisis (Anexo II). El primer análisis (IA1) aplicó la rúbrica sin contextualización adicional. El segundo (IA2) añadió información evolutiva relevante para niños de tres años –etapa previa al esquema gráfico, tolerancia a omisiones anatómicas y énfasis en la intención por encima de la precisión técnica–. Tras finalizar IA2, se efectuó un tercer análisis (IA3) que repitió el proceso, pero solicitó explícitamente no focalizar la valoración en la figura humana. Finalmente, se generó IA4, una réplica exacta de IA2 lanzada desde otra cuenta independiente (Universidad de Salamanca), con el objetivo de comprobar la estabilidad interinstancia del modelo. En los cuatro casos, GPT-4o procesó cada imagen, asignó las doce puntuaciones, y añadió un comentario opcional, todo ello volcado automáticamente en la plantilla Excel usada también por los evaluadores humanos. De este modo fue posible comparar la consistencia interna y la convergencia entre las valoraciones humanas y las generadas por la inteligencia artificial bajo condiciones metodológicamente equivalentes.

## 2.5. Análisis estadístico

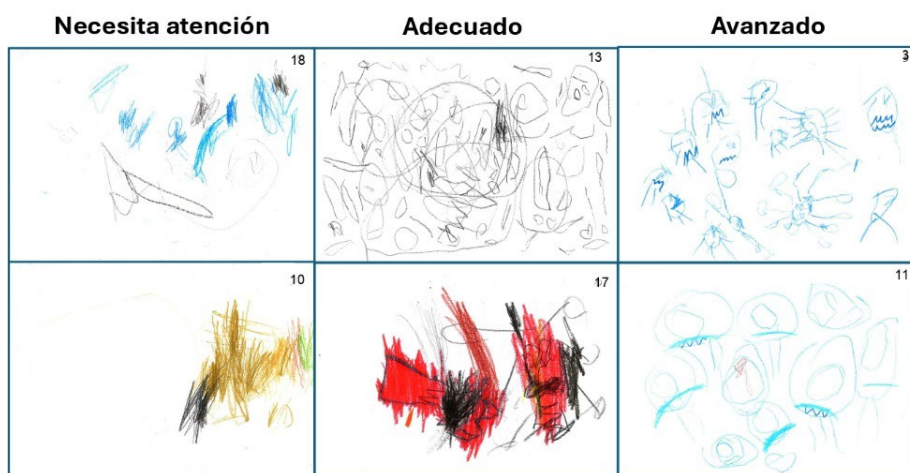
Los datos obtenidos fueron procesados y analizados mediante el software estadístico Jamovi (versión 2.6.44), basado en R. Con el propósito de comparar las puntuaciones asignadas por los evaluadores humanos (E1 a E5) y por el modelo de inteligencia artificial GPT-4o (IA1 a IA4), se empleó la prueba no paramétrica de Wilcoxon para muestras apareadas, adecuada para distribuciones no necesariamente normales y para datos ordinales derivados de rúbricas cualitativas. Asimismo, se utilizó el coeficiente de correlación de Pearson con el fin de examinar el grado de asociación lineal entre las valoraciones. Este análisis permitió detectar posibles concordancias o discrepancias sistemáticas entre ambos sistemas de evaluación en cada una de las dimensiones analizadas: desarrollo motor fino, desarrollo cognitivo y de representación, desarrollo de indicadores emocionales y dimensión social/ autoimagen.

Adicionalmente, se estimó la consistencia interna de los sistemas de evaluación mediante el cálculo del coeficiente alfa de Cronbach. Esta medida se empleó para valorar la homogeneidad de cada una de las dimensiones estimadas por los expertos y la IA. La interpretación de este coeficiente permite aproximarse a la fiabilidad del instrumento de evaluación, entendida como la estabilidad interna en la aplicación de los criterios sobre las producciones analizadas. Cuando la fiabilidad de las dimensiones era alta, se realizó un análisis factorial exploratorio de las cuatro dimensiones, empleando factorización de ejes principales como método de extracción y una rotación oblimin, lo que permitió estimar la ponderación de cada dimensión en el desarrollo gráfico-expresivo.

### 3. RESULTADOS

#### 3.1. Análisis de los dibujos por dimensiones y fuentes de evaluación

El examen de las 22 representaciones de microorganismos, calificadas por cinco jueces expertos (E1–E5) y por cuatro instancias independientes de GPT-4o (IA1–IA4), evidenció una distribución amplia y equilibrada de los tres niveles de logro («Necesita atención», «Adecuado» y «Avanzado») en los catorce indicadores que conforman las cuatro dimensiones de la rúbrica: desarrollo motriz fino (DMF), desarrollo cognitivo y de representación (DCR), indicadores emocionales (DIE) y dimensión social/autoimagen (DS). Este patrón evidencia la capacidad discriminativa del instrumento (Figura 1).



**Figura 1.** Ejemplo de dibujos proporcionados por los alumnos

#### 3.2. Descripción cualitativa de los niveles

Los dibujos ubicados en «Necesita atención» se caracterizaron por grafismos esquemáticos, paleta cromática restringida y omisión de detalles morfológicos; la localización espacial fue, en ocasiones, periférica o descentrada, lo que denota un aprovechamiento limitado de los recursos expresivos. En el nivel «Adecuado» se incluyeron los que presentaban siluetas microbianas reconocibles con rasgos internos incipientes y un predominio de tonos oscuros (negro y negro-rojo), indicio de la connotación negativa atribuida al concepto microbiano. Las producciones catalogadas como «Avanzado» incorporaron estructuras morfológicas diferenciadas, mayor riqueza cromática e intensificación gestual (bocas dentadas, ojos airados), indicadores de una comprensión más elaborada del estímulo y de una mayor madurez gráfica.

Estas características fueron recogidas parcialmente por los juicios generados por GPT-4o, como se aprecia en los comentarios literales que la IA emitió, por ejemplo, para los seis dibujos mostrados en la Figura 1:

- Dibujo 18 («Necesita atención»): “El dibujo revela control gráfico y trazo seguro, aunque sin riqueza simbólica ni elementos de representación adicionales. Se destaca por su ejecución técnica”.

- Dibujo 10 («Necesita atención»): “El dibujo se organiza en torno a una figura central simple, con trazos controlados y uso variado de color. No hay intención narrativa ni

elementos sociales, pero se observa un nivel gráfico coherente con un desarrollo preesquemático intermedio. Puede beneficiarse de consignas que estimulen la representación de múltiples entidades o interacciones”.

- Dibujo 13 («Adecuado»): “El dibujo presenta una estructura básica estable, con formas reconocibles y aplicación cromática consistente. Aunque no hay elementos narrativos ni sociales, la organización general indica un desarrollo gráfico funcional dentro de los parámetros de la etapa previa al esquema gráfico”.

- Dibujo 17 («Adecuado»): “El dibujo muestra un trazo seguro y control del espacio, pero limitado en su riqueza en la representación. La escasa variedad cromática y ausencia de narrativa o relaciones espaciales sugieren un estadio inicial dentro de la etapa previa al esquema gráfico. Requiere seguimiento para observar posibles progresiones”.

- Dibujo 3 («Avanzado»): “El dibujo revela una representación organizada con indicios de forma orgánica y movilidad. La diversidad de trazos y colores indica exploración simbólica en desarrollo. El estilo confirma avances en control gráfico y configuración estructural”.

- Dibujo 11 («Avanzado»): “Este dibujo destaca por su madurez gráfica, con formas diferenciadas, uso preciso del espacio y un manejo cromático expresivo. Aunque no incluye interacción social ni figuras humanas, el nivel de estructuración visual indica un avance en la representación notable dentro del marco previo al esquema gráfico”.

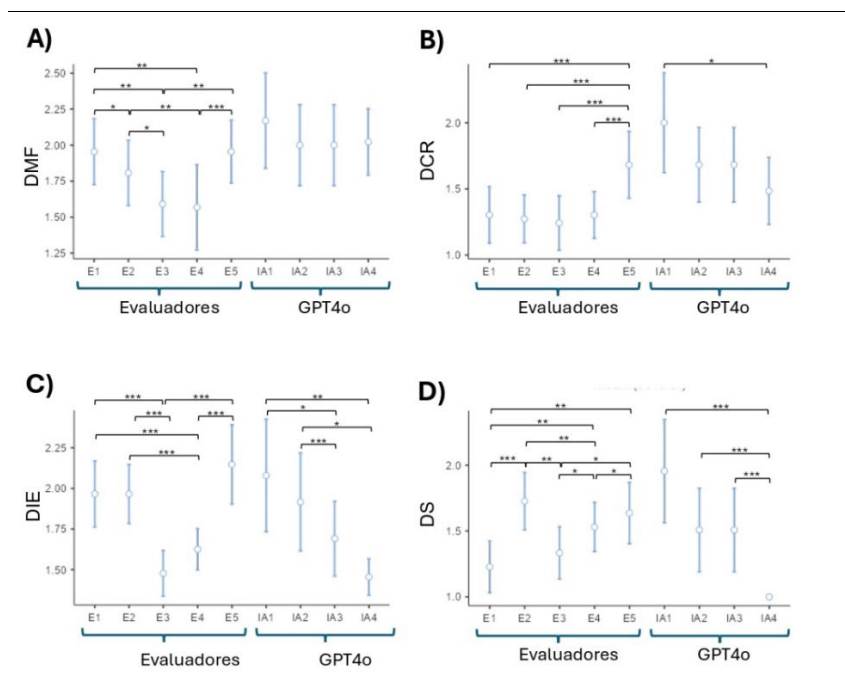
Los comentarios de GPT-4o convergen con las observaciones expertas en variables formales –trazo, cromaticidad y uso del espacio–, pero divergen en la identificación simbólica: la IA atribuyó expresiones antropomórficas inexistentes a los dibujos 3 y 11, lo que revela ciertas limitaciones interpretativas.

### 3.3. Comparación cuantitativa entre expertos e IA

La evaluación de la concordancia cuantitativa entre las valoraciones de los cinco expertos y entre las producidas por las cuatro instancias de GPT-4o mediante la prueba de rangos con signo de Wilcoxon (datos apareados), mostró desigual consistencia interna entre ambos sistemas de evaluación. El contraste reveló discrepancias estadísticamente significativas en las cuatro dimensiones cuando se tomó como referencia el conjunto de puntuaciones humanas ( $p < 0,05$ ). En las puntuaciones generadas por la IA la divergencia fue menor: no se observaron diferencias significativas en las valoraciones proporcionadas por la IA para el desarrollo motor fino (DMF), mientras que las dimensiones desarrollo cognitivo y de representación (DCR), indicadores emocionales (DIE) y dimensión social/autoimagen (DS) sí mostraron diferencias, aunque con menos frecuencia que entre expertos (Figura 2). La dispersión interexperto fue mayor en las puntuaciones humanas –especialmente en DMF, DIE y DS– lo que apunta a una aplicación heterogénea de los criterios de la rúbrica. Por el contrario, las cuatro instancias de GPT-4o ofrecieron estimaciones internas más homogéneas y estrechas, si bien mantuvieron cierta variabilidad en DIE y DS, dimensiones que exigen mayor interpretación afectiva y contextual.

Estos hallazgos indican que el valor absoluto asignado por la rúbrica depende en gran medida del evaluador, lo que limita su utilidad para comparaciones interexpertos sin procedimientos adicionales de calibración. La consistencia relativa observada en la IA sugiere que GPT-4o aplica la rúbrica de manera estable, pero su sensibilidad interpretativa sigue siendo inferior en los indicadores emocionales y sociales; por ello, la

rúbrica resulta más adecuada para el seguimiento intraevaluador y requiere supervisión especializada cuando se emplea como apoyo automatizado.



**Figura 2.** Valores medios obtenidos para cada una de las dimensiones por cada evaluador (E1 a E5) y por cada uno de los modelos de inteligencia artificial (IA1 a IA4). Las barras de error representan el intervalo de confianza al 95 %; los asteriscos indican significación del contraste Wilcoxon (\*  $p \leq .05$ ; \*\*  $p \leq .01$ ; \*\*\*  $p \leq .001$ ). Dimensiones: A) Desarrollo Motriz Fino (DMF), B) Desarrollo Cognitivo y de Representación (DCR), C) Desarrollo de Indicadores Emocionales (DIE), D) y Dimensión Social / Autoimagen (DS).

### 3.4. Consistencia interna y fiabilidad de los sistemas de evaluación

La coherencia entre jueces se examinó, en primer lugar, mediante la correlación de Pearson aplicada a las puntuaciones de cada dimensión. Las cinco rúbricas humanas (E1–E5) mostraron asociaciones positivas y elevadas en las cuatro escalas ( $r = 0,62-0,88$ ,  $p < 0,001$ ), lo que indica que, aun cuando difieren en el valor absoluto asignado, los evaluadores coinciden al ordenar los dibujos según su calidad relativa. En contraste, las calificaciones emitidas por las cuatro instancias GPT-4o presentaron correlaciones moderadas o nulas ( $r = -0,25-0,41$ ) y, en el 58 % de los pares IA-IA, la relación careció de significación estadística. Las divergencias fueron especialmente acusadas en desarrollo cognitivo y de representación (DCR) y en indicadores emocionales (DIE), donde aparecieron incluso correlaciones negativas entre algunas instancias de IA y los jueces humanos.



$\alpha$ Cronbach	DMF	DCR	DIE	DS
E	0,942	0,893	0,835	0,852
IA	0,434	-0,526	0,217	-0,203

**Tabla 1.** Análisis de fiabilidad mediante alfa de Cronbach de las evaluaciones realizadas por humanos (E) y las realizadas mediante ChatGPT4o (IA)

### 3.5. Análisis factorial exploratorio e Índice de Desarrollo Gráfico-Expresivo (IDGE)

Dada la alta fiabilidad interna observada en las evaluaciones humanas, se promediaron las puntuaciones de los cinco jueces para cada una de las cuatro dimensiones y se sometieron a un análisis factorial exploratorio mediante factorización de ejes principales con rotación oblimin. El índice de adecuación muestral de Kaiser-Meyer-Olkin fue satisfactorio ( $KMO = 0,84$ ) y la prueba de esfericidad de Bartlett resultó significativa ( $\chi^2 = 48,3$ ,  $p < 0,001$ ), confirmando la idoneidad de la estructura correlacional.

La solución de un solo factor mostró un ajuste excelente ( $\chi^2 = 0,07$ ,  $p = 0,97$ ) y explicó el 83% de la varianza común. Este factor –denominado Índice de Desarrollo Gráfico-Expresivo (IDGE)– recibió cargas muy elevadas de DMF (0,99), DCR (0,96) y DS (0,92), y una carga moderada de DIE (0,48) (Tabla 2), lo que sugiere que los indicadores motriz, cognitivo, social y de representación comparten un núcleo latente robusto, mientras que la dimensión emocional conserva cierta especificidad.

Factor	Carga factorial	Varianza única	KMO
DMF	0,986	0,028	0,718
DCR	0,959	0,079	0,765
DIE	0,480	0,769	0,979
DS	0,923	0,147	0,883

**Tabla 2.** Análisis factorial exploratorio realizado con los promedios de las puntuaciones de los evaluadores humanos (E1 a E5) para cada dimensión (Desarrollo motriz fino DMF, Desarrollo cognitivo y de representación DCR, Desarrollo de indicadores emocionales DIE, Dimensión social/autoimagen DS). Para cada factor se incluye la carga factorial, la varianza única explicada y el test Kaiser-Meyer-Olkin (KMO) como medida de adecuación de los datos

Partiendo de las cargas factoriales estandarizadas y del método de regresión para estimar las puntuaciones de factor, se propone ponderar el IDGE asignando un 30% a DMF, 30% a DCR, 30% a DS y 10% a DIE. Ello permite calcular un índice sintético de desarrollo gráfico-expresivo para cada uno de los 22 dibujos según cada evaluador.

Las puntuaciones de IDGE obtenidas por los jueces difirieron de forma significativa entre sí (Wilcoxon pareado,  $p < 0,05$ ), lo que ratifica la imposibilidad de emplear el índice como valor absoluto intercambiable entre evaluadores. No obstante, la consistencia interna del IDGE resultó excelente ( $\alpha = 0,95$ ), superando la fiabilidad de cualquiera de las dimensiones por separado. Por tanto, dentro de un mismo evaluador, el IDGE constituye una medida robusta para ordenar a los alumnos según su nivel de desarrollo gráfico-expresivo, si bien requiere calibración rigurosa para comparaciones entre jueces.

## 4. DISCUSIÓN

El dibujo infantil representa una vía de expresión integral que articula dimensiones cognitivas, emocionales y sociales. Elementos como el uso del color, la intensidad del trazo, la proporción de figuras y la inclusión de detalles anatómicos son indicativos de aspectos del desarrollo gráfico expresivo. Estos aspectos, incluidos en distintos trabajos previos sobre análisis de dibujos en la etapa de Educación Infantil (Farokhi y Hashemi, 2011; Lowenfeld y Brittain, 1987; Moreno, 2022; Zlateva, 2019), han sido integrados en las distintas dimensiones analizadas con la rúbrica permitiendo así concretar esas variables y convertirlas en indicadores observables.

El análisis de dibujos realizados por alumnado de 3-4 años sobre microorganismos, a partir de la aplicación de una rúbrica, tanto por expertos humanos como por el modelo GPT-4o, pone de manifiesto un elevado grado de agresividad en las representaciones gráficas de este contenido. Esto se observa tanto en las producciones gráficas clasificadas como adecuadas, caracterizadas por un uso extendido de combinaciones de colores oscuros, principalmente negro y rojo (Moreno, 2022; Zlateva, 2019), como en las avanzadas, que incluyen representaciones antropomórficas con rostros agresivos (Farokhi y Hashemi, 2011). De acuerdo con estos autores, la inclusión de estas variables expresivas podría reflejar estados emocionales y experiencias personales de los participantes sobre los microorganismos difícilmente detectables por otros medios. El análisis de los dibujos muestra que desde comienzos de la etapa de Educación Infantil los alumnos tienen una concepción negativa sobre los microorganismos, situación que ya se había observado en edades superiores (Ballesteros et al., 2018; Byrne, 2011; Domínguez et al., 2018; Faccio et al., 2013). Esta circunstancia es relevante para el profesorado de esta etapa que, a través de los dibujos y el uso de la rúbrica para su análisis, puede confirmar si esta concepción está presente en sus alumnos e intervenir tempranamente sobre ella.

En lo metodológico, los resultados obtenidos reflejan que las evaluaciones realizadas por los expertos humanos presentan una alta dispersión, pero una elevada fiabilidad y consistencia interna, en contraste con las generadas por inteligencia artificial, que evidenciaron una menor dispersión, pero baja consistencia. Este hallazgo resalta la capacidad del juicio profesional para captar matices expresivos vinculados al desarrollo infantil que la IA, en su estado actual, no logra interpretar adecuadamente. Los evaluadores humanos, apoyándose en distintos marcos teóricos sobre el desarrollo gráfico infantil (Burt, 1921; Luquet, 1981; Rinaldi, 2006; Wright, 2007), identificaron patrones evolutivos propios del dibujo infantil, reconociendo elementos simbólicos y de representación que la IA no interpretó con precisión. Además, las correlaciones estadísticas entre evaluadores humanos fueron significativamente más altas que aquellas obtenidas con la IA, lo cual refuerza su mayor fiabilidad. Este bajo nivel de fiabilidad interna en las puntuaciones generadas por IA pone en cuestión la estabilidad y coherencia del sistema automatizado. Resulta especialmente revelador que, utilizando el mismo prompt y la misma información de entrada, dos usuarios distintos (IA2 e IA4) hayan obtenido evaluaciones divergentes por parte de ChatGPT. Este hallazgo refuerza la percepción de que, en su estado actual, los sistemas de inteligencia artificial no garantizan resultados consistentes ni reproducibles en contextos de evaluación cualitativa, comprometiendo así su aplicabilidad como herramienta de apoyo en entornos académicos donde la objetividad y la fiabilidad son esenciales. No obstante, la IA presenta ventajas operativas como la velocidad, replicabilidad y capacidad de procesar grandes volúmenes

de datos. Además, los comentarios que elabora la IA pueden guiar la evaluación de un juicio experto. Estas características podrían aprovecharse en tareas preliminares o como apoyo al análisis profesional, siempre bajo supervisión experta.

El análisis factorial realizado con las puntuaciones de los evaluadores humanos pone de manifiesto la existencia de un factor latente que respalda la validez del instrumento diseñado para evaluar la progresión en los dibujos infantiles, lo que ha permitido construir el Índice de Desarrollo Gráfico-Expresivo. Este índice no es aplicable para comparar valoraciones emitidas por diferentes expertos humanos, pero sí permite que un mismo evaluador establezca un orden de desarrollo entre distintos alumnos. Así, el instrumento facilita que un docente de Educación Infantil analice las representaciones gráficas de su grupo de alumnos, identificando qué niños requieren mayor atención y cuáles presentan un nivel de desarrollo más avanzado. La rúbrica desarrollada resulta especialmente relevante si se considera que, actualmente, no existen herramientas cualitativas estandarizadas que permitan codificar de manera sistemática las producciones gráficas infantiles, lo que limita considerablemente su interpretación objetiva (Campbell y Cunningham, 2025). La propuesta desarrollada en este trabajo contribuye a superar esta carencia al ofrecer un marco estructurado y fiable para la evaluación del desarrollo gráfico-expresivo en la infancia.

En síntesis, si bien la IA puede ser útil como herramienta complementaria, la interpretación profunda de los dibujos infantiles continúa dependiendo del juicio humano, ya que los aspectos incluidos en la rúbrica requieren de una interpretación contextualizada. Futuras investigaciones deberán explorar mejoras en los modelos de IA para aumentar su capacidad interpretativa, sin comprometer la calidad del análisis pedagógico y emocional que exige este tipo de evaluación.

El estudio planteado presenta ciertas limitaciones que hay que considerar de cara a su generalización. En primer lugar, la muestra –22 alumnos y 5 evaluadores– es reducida; replicar el procedimiento con un número mayor de participantes y distintos contextos socioculturales reforzaría la validez externa. En segundo lugar, la fidelidad de la rúbrica, aunque alta, debería contrastarse con autopercepciones de los niños mediante entrevistas o técnicas proyectivas, a fin de triangular datos y consolidar la validez de contenido. Por último, se analizó un único tipo de consigna (“dibuja un microorganismo”); futuras investigaciones deberían comprobar si la estructura factorial se mantiene con tareas gráficas de otra naturaleza.

## 5. CONCLUSIONES

La rúbrica desarrollada en esta investigación constituye una herramienta válida y fiable para la evaluación del dibujo infantil cuando es aplicada por evaluadores humanos, como lo evidencian los elevados coeficientes de fiabilidad y la alta concordancia entre evaluadores. Su diseño, fundamentado en la literatura científica y ajustado a criterios del desarrollo infantil, permite valorar de forma sistemática distintas dimensiones –motrices, cognitivas, emocionales y sociales– en alumnado a partir de los 3 años.

Según los resultados obtenidos mediante el análisis factorial exploratorio de las puntuaciones emitidas por evaluadores humanos, esta rúbrica puede resultar especialmente útil en las aulas de Educación Infantil, ya que ofrece a los docentes una herramienta estructurada para valorar el desempeño gráfico de su alumnado y evaluar sus

concepciones tempranas sobre distintos contenidos. En cambio, las evaluaciones generadas por el modelo GPT-4o en su estado actual, si bien se llevaron a cabo con los mismos criterios y rúbrica, no alcanzan niveles de fiabilidad aceptables. Los análisis revelaron correlaciones débiles y una baja consistencia interna, lo que compromete su aplicabilidad directa en procesos de diagnóstico o intervención. Por tanto, no se recomienda su uso como herramienta de evaluación del desarrollo infantil en contextos educativos reales.

No obstante, el uso de inteligencia artificial podría ofrecer cierto valor como recurso complementario bajo supervisión profesional. Por ejemplo, podría facilitar tareas de cribado inicial o análisis exploratorio, y ayudar al profesorado a identificar concepciones erróneas tempranas –como la visión agresiva o peligrosa de los microorganismos observada en algunos dibujos–, orientando así la planificación de acciones pedagógicas correctivas.

## AGRADECIMIENTOS

Este trabajo ha sido financiado por la subvención PID2022-139684NB-I00, concedida por el Ministerio de Ciencia, Innovación y Universidades (MCIN)/Agencia Estatal de Investigación (AEI) (10.13039/501100011033), y cofinanciado por el Fondo Europeo de Desarrollo Regional (FEDER), 'Una manera de hacer Europa'.

## BIBLIOGRAFÍA

- Aparicio Gómez, W.O. (2023). La inteligencia artificial y su incidencia en la educación: Transformando el aprendizaje para el siglo XXI. *Revista Internacional de Pedagogía e Innovación Educativa*, 3, 217-229. <https://doi.org/10.51660/ripie.v3i2.133>
- Asociación Médica Mundial (2024, 23 de octubre). *Resumen de la Asamblea General de la Asociación Médica Mundial 2024*. <https://www.wma.net/es/news-post/resumen-de-la-asamblea-general-de-la-asociacion-medica-mundial-2024/>
- Ayuso del Puerto, D. y Gutiérrez Esteban, P. (2022). La inteligencia artificial como recurso educativo durante la formación inicial del profesorado. *RIED-Revista Iberoamericana de Educación a Distancia*, 25(2), 347-362. <https://doi.org/10.5944/RIED.25.2.32332>
- Ballesteros, M.I., Paños Martínez, E. y Ruiz-Gallardo, J.R. (2018). Los microorganismos en la educación primaria: Ideas de los alumnos de 8 a 11 años e influencia de los libros de texto. *Enseñanza de las Ciencias*, 36(1), 79-98. <https://doi.org/10.5565/rev/ensciencias.2274>
- Bazzari, A.H. y Bazzari, F.H. (2024). Assessing the ability of GPT-4o to visually recognize medications and provide patient education. *Scientific Reports*, 14(1), 26749. <https://doi.org/10.1038/S41598-024-78577-Y>
- Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y.T., Li, Y., Lundberg, S., Nori, H., Palangi, H., Ribeiro, M.T. y Zhang, Y. (2023). *Sparks of artificial general intelligence: Early experiments with GPT-4*. arXiv preprint. <https://arxiv.org/abs/2303.12712>
- Burt, C.L. (1921). *Mental and scholastic tests*. P.S. King.
- Byrne, J. (2011). Models of micro-organisms: Children's knowledge and understanding of micro-organisms from 7 to 14 years old. *International Journal of Science Education*, 33(14), 1.927-1.961. <https://doi.org/10.1080/09500693.2010.536999>
- Campbell, K. y Cunningham, A.E. (2025). Examining young children's transitions from drawing into early writing. *Early Childhood Research Quarterly*, 72, 317-327. <https://doi.org/10.1016/j.ecresq.2025.04.007>
- Cherney, I.D., Seiwert, C.S., Dickey, T.M. y Flichtbeil, J.D. (2006). Children's drawings: A mirror to their minds. *Educational Psychology*, 26(1), 127-142. <https://doi.org/10.1080/01443410500344167>
- Cox, M. (2005). *The Pictorial World of the Child*. Cambridge University Press.

- Domínguez, C.R.C., Leporo, N., de Franco, M.T., Inglez, G.C., Gonçalves, V.M. y Bizerra, A.F. (2018). Learning about microorganisms in childhood: Four to six year old children's voice in kindergartens and museums. *Revista Brasileira de Pesquisa em Educação em Ciências*, 1(2), 1-25. <https://periodicos.ufmg.br/index.php/rbpec/article/view/4583/2988>
- Faccio, E., Costa, N., Losasso, C., Cappa, V., Mantovani, C., Cibin, V., Andrighetto, I. y Ricci, A. (2013). What programs work to promote health for children? Exploring beliefs on microorganisms and on food safety control behavior in primary schools. *Food Control*, 33(2), 320-329. <https://doi.org/10.1016/j.foodcont.2013.03.005>
- Farokhi, M. y Hashemi, M. (2011). The analysis of children's drawings: Social, emotional, physical, and psychological aspects. *Procedia - Social and Behavioral Sciences*, 30, 2.219-2.224. <https://doi.org/10.1016/j.sbspro.2011.10.433>
- García-Peña, V.R., Mora-Marcillo, A.B. y Ávila-Ramírez, J.A. (2020). La inteligencia artificial en la educación. *Dominio de las Ciencias*, 6(3), 648-666.
- Granda Dávila, M.F., Muncha Cofre, I.J., Guamanquispe Rosero, F.V. y Jácome Noroña, J.H. (2024). Inteligencia artificial: Ventajas y desventajas de su uso en el proceso de enseñanza-aprendizaje. *MENTOR Revista de Investigación Educativa y Deportiva*, 3(7), 202-224. <https://doi.org/10.56200/mried.v3i7.7081>
- Jamovi Project (2024). *Jamovi (Version 2.6.44) [Computer software]*. <https://www.jamovi.org>
- Lowenfeld, V. y Brittain, W.L. (1987). *Creative and mental growth (8th ed.)*. Prentice Hall.
- Luquet, G.H. (1981). *El dibujo infantil*. Editorial Médica y Técnica SA.
- Maestre Castro, A.B. (2010). El dibujo en la escuela. *Innovación y Experiencia Educativa*, 26, 1-12. [https://archivos.csif.es/archivos/andalucia/ensenanza/revistas/csicsif/revista/pdf/Numero\\_26/ANA\\_BELEN\\_MAESTRE\\_CASTRO\\_01.pdf](https://archivos.csif.es/archivos/andalucia/ensenanza/revistas/csicsif/revista/pdf/Numero_26/ANA_BELEN_MAESTRE_CASTRO_01.pdf)
- Malchiodi, C.A. (2012). *Understanding children's drawings*. Guilford Publications.
- Molina-Jiménez, A. (2015). El dibujo infantil: Trazos, colores e historias que nos hacen reflexionar y aprender. *Revista Electrónica Educare*, 19(1), 167-182. <https://doi.org/10.15359/ree.19-1.10>
- Moreno González, A. (2022). *La interpretación del dibujo infantil*. Octaedro.
- OpenAI (2024). *GPT-4o [Large language model]*. <https://openai.com/gpt-4o>
- Quiroga Méndez, P. (2007). Etapas de desarrollo del dibujo infantil, entre el constructivismo y el ambientalismo. *Papeles Salmantinos de Educación*, 9, 255-282. <https://doi.org/10.36576/summa.29547>
- R Core Team (2024). *R: A language and environment for statistical computing [Computer software]*. R Foundation for Statistical Computing. <https://www.r-project.org/>
- Rathje, S., Dan-Mircea, M., Lia, S., Raja, M., Robertson, C.E. y Van Bavel, J.J. (2024). GPT is an effective tool for multilingual psychological text analysis. *Proceedings of the National Academy of Sciences of the United States of America*, 121(34), e2308950121. <https://doi.org/10.1073/pnas.2308950121>
- Rinaldi, C. (2006). *In dialogue with Reggio Emilia: Listening, researching and learning*. Routledge.
- Rouhiainen, L. (2018). *Inteligencia artificial: 101 cosas que debes saber hoy sobre nuestro futuro*. Alienta.
- Shah, U., Khan, N., Alzubaidi, M., Agus, M. y Househ, M. (2025). ArtInsight: A multimodal AI framework for interpreting children's drawings and enhancing emotional understanding. *Studies in Health Technology and Informatics*, 327, 808-812. <https://doi.org/10.3233/SHTI250471>
- Vygotsky, L.S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Wright, S. (2007). *Children, meaning-making and the arts*. Pearson Education.
- Zlateva, A. (2019). *How to read children's drawings*. University of Malta. <https://www.um.edu.mt/library/oar/handle/123456789/48957>

## ANEXO 1. RÚBRICA

Esta rúbrica apoya la observación y evaluación cualitativa de dibujos realizados por niñas y niños de 3 a 6 años. Permite identificar logros y áreas que requieren seguimiento en cuatro dimensiones del desarrollo: motriz, cognitivo y de representación, emocional y social. Cada indicador se califica con una escala tricotómica: 1 = Necesita atención • 2 = Adecuado (rango típico) • 3 = Avanzado.

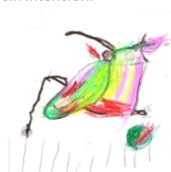
**Nota:** La rúbrica se configuró con muestras de 3-6 años. Para 6-8 años utilice la «Tabla de ajustes por edad» del Anexo A antes de interpretar las puntuaciones. Definiciones operativas clave:

- **Fondo:** Se consideran fondo elementos ambientales como cielo, suelo, mobiliario, paisaje o decorado que no son figuras principales.
- **Omisión o distorsión de partes del cuerpo:** Antes de los 5 años la ausencia de cuello no se clasifica como distorsión grave.

### 1. Desarrollo motriz fino (DMF)

Indicador	1 – Necesita atención	2 – Adecuado	3 – Avanzado
Control de trazo	Trazos temblorosos, discontinuos o sin dirección controlada.	Trazos rectos y curvas aceptables; algunas desviaciones menores.	Trazos seguros, continuos y bien dirigidos.
Presión / continuidad del trazo	Presión muy desigual; cortes frecuentes que denotan falta de control.	Presión moderadamente uniforme; pocos cortes en el trazo.	Presión constante y adecuada, sin interrupciones perceptibles.
Formas básicas	Garabatos sin formas geométricas reconocibles.	Formas simples (círculos, cuadrados) aunque irregulares.	Uso variado de formas claras para componer figuras significativas.
Uso del espacio	El dibujo se concentra en un rincón o disperso sin orden.	Distribución parcial; quedan áreas grandes vacías sin intención.	Composición equilibrada que aprovecha la mayor parte de la hoja.

Ejemplos:




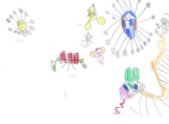

### 2. Desarrollo cognitivo y de representación (DCR)

Indicador	1 – Necesita atención	2 – Adecuado	3 – Avanzado
Detalle y complejidad	≤2 elementos simples sin detalles.	Figura reconocible con algunos detalles (cabello, ropa).	Figuras completas y múltiples detalles (rostro, fondo, accesorios).
Relaciones espaciales	Elementos amontonados o sin relación espacial clara.	Alguna relación (encima, al lado) pero inconsistente.	Relaciones espaciales coherentes (delante/detrás, cerca/lejos).
Narrativa visual	Escena sin acción; elementos inconexos.	Indicios de acción o de relación temática sencilla.	Escena cuenta una historia clara con secuencia o interacciones.




Ejemplos:



### 3. Indicadores emocionales (DIE)

Indicador	1 – Necesita atención	2 – Adecuado	3 – Avanzado
Uso del color	1–2 colores sin diferenciación entre partes.	≥3 colores con algún criterio, aunque inconsistente.	Colores asignados a elementos específicos o con intención emocional.
Intensidad cromática	Color tenue o sin relleno.	Aplicación moderada; intensidad desigual.	Colores saturados o sombreados que enfatizan emoción/elemento.
Tamaño y posición de la figura humana	Figura desproporcionada o en el borde.	Figura centrada con tamaño relativamente adecuado.	Figura proporcionada y situada coherentemente con la escena.
Omisión o distorsión de partes del cuerpo	Faltan partes esenciales (cabeza, brazos).	Figura incompleta o partes simplificadas; cuello ausente aceptable <5 años.	Figura completa con detalles anatómicos pertinentes.
Ejemplos:			

### 4. Social / auto-imagen (DS)

Indicador	1 – Necesita atención	2 – Adecuado	3 – Avanzado
Inclusión de figuras significativas	Solo 1 figura o ninguna.	2 figuras presentes sin interacción evidente.	≥3 figuras con interacción (contacto, acción compartida).
Interacción explícita	Figuras aisladas sin contacto visual ni actividad común.	Alguna orientación mutua o proximidad sin acción clara.	Miradas, gestos o acciones conjuntas claramente representadas.
Elementos del fondo	Sin fondo o solo color sólido.	Fondo básico con 1–2 elementos (cielo, sue	Fondo complejo con ≥3 elementos (árboles, casa, mobiliario).
Ejemplos:			

#### Procedimiento de puntuación

1. Evalúe cada indicador por separado y asigne un valor de 1 a 3.
2. Calcule la media parcial de cada dimensión.
3. Obtenga la media global ponderada con los pesos:
  - Motriz 30 % • Cognitivo 30 % • Emocional 10 % • Social 30 %.
4. Registre la varianza entre dibujos para detectar incoherencias intra-niño/a.

Esta rúbrica no constituye un instrumento diagnóstico clínico. Los resultados señalan áreas que pueden requerir observación adicional. Considere las variaciones culturales (por ejemplo, el simbolismo del color o la representación del espacio) antes de inferir conclusiones emocionales o sociales.

#### Anexo A. Tabla de ajustes por edad (6 – 8 años)

Dimensión	Nivel 2 (Adecuado) – Ajuste	Nivel 3 (Avanzado) – Ajuste
DMF	Trazos continuos y seguros sin desviaciones significativas; presión uniforme.	Control fino preciso; incorporación de perspectiva simple y detalles anatómicos.
DCR	Escena coherente con relaciones espaciales consistentes y varios detalles.	Narrativa compleja con múltiples acciones secuenciales y perspectiva básica.
DIE	Colores asignados temáticamente; saturación e intensidad apropiadas.	Uso de color simbólico/emocional coherente y sombreado para volumen.
DS	Interacción básica (mirada o proximidad) entre al menos dos figuras.	Interacción compleja con roles diferenciados y cooperación explícita.

## ANEXO 2. DETALLE DE LAS INSTRUCCIONES (PROMPTS) UTILIZADAS EN CHATGPT4O

El primer análisis automatizado (IA1) se efectuó con el siguiente prompt:

Eres un especialista en psicología del desarrollo infantil con experiencia en análisis gráfico infantil. Vas a analizar dibujos realizados por niños de 3 años que respondieron a la consigna: **\*\*“Dibuja un microorganismo”\*\***.

**OBJETIVO:**

Aplicar una rúbrica científica estructurada en cuatro dimensiones del desarrollo (motriz, cognitivo, emocional y social), cada una con indicadores claros y escala de 1 a 3 puntos. Evaluar cada dibujo, justificar brevemente las puntuaciones si es posible, y registrar los resultados en una plantilla Excel (una fila por dibujo).

**MATERIAL:**

Rúbrica oficial basada en literatura científica (ya cargada).  
Plantilla Excel estructurada para recoger los resultados (ya cargada).

#### INSTRUCCIONES:

Al recibir cada imagen de dibujo:

Analízala con base en la rúbrica “Rubrica\_Analisis\_Dibujos\_Infantiles\_v4.pdf”.

Evalúa los 12 indicadores divididos en 4 dimensiones:

**\*\*Desarrollo Motriz Fino (DMF):\*\*** Control del trazo, presión, formas básicas, uso del espacio.

**\*\*Desarrollo Cognitivo Representacional (DCR):\*\*** Detalle y complejidad, relaciones espaciales, narrativa visual.

**\*\*Indicadores Emocionales (DIE):\*\*** Uso del color, intensidad cromática, figura humana.

**\*\*Dimensión Social / Autoimagen (DS):\*\*** Inclusión de figuras significativas, interacción explícita, elementos del fondo.

Asigna una puntuación de 1 a 3 a cada indicador, conforme a la rúbrica.

Calcula la media de cada dimensión.

Redondea la media final a un decimal.

Opcionalmente, incluye un comentario breve sobre el dibujo (elementos clave, color, intención).

Registra en la plantilla Excel:

ID del dibujo (ej. Dibujo\_01)

Puntuaciones individuales (12 indicadores)

Media de cada dimensión

Media global ponderada

Comentario opcional

Cuando te proporcione el primer dibujo, realiza el análisis y genera la fila correspondiente para la Excel.

En el segundo análisis (IA2) se incorporó información adicional sobre la edad de los niños, ajustando el análisis a su etapa evolutiva preesquemática. Se añadió:

#### CONTEXTO:

Los niños tienen 3 años.

A esta edad es habitual la omisión de cuello, extremidades o proporciones; no se penaliza automáticamente.

Ten en cuenta la posible abstracción y libertad expresiva típica de la etapa preesquemática.

Considera la intención, no sólo la precisión técnica.