DISEÑO Y PILOTAJE DE PRUEBA PARA FUNCIONES VISOCONSTRUCCIONALES: TEST DE ORIÓN

Andrea Nicole **Alarcón Bello /** *analarconb@unal.edu.co*
Nila Fernanda **Amaya Melo /** *nfamayam@unal.edu.co*
Laura Alejandra **Correa González /** *laacorreago@unal.edu.co*
Julián Andrés **Lapeira Rincón /** *jalapeirar@unal.edu.co*

Estudiantes del pregrado en Psicología
Universidad Nacional de Colombia, sede Bogotá

Palabras clave: *Funciones visoconstruccionales, funciones visoespaciales, orientación espacial, configuración holística.*

RESUMEN

El presente artículo presenta el diseño de una prueba, cuyo objetivo era medir las funciones visocontruccionales (compuestas por orientación espacial, actividad perceptual, capacidad motriz y funciones ejecutivas). Para ello, se tomaron como base los procedimientos de copia propuestos por Grossi y Angelini (1991, citados en Montañés et al., 2011) para la elaboración de los ítem, los cuales se aplicaron a 65 personas entre los 16 y los 100 años. Se realizó análisis psicométrico para evidencias de validez (análisis de ítems, evidencia basada en la relación con otras variables y juicio de expertos) y confiabilidad (Alfa de cronbach). Los resultados mostraron que la consistencia interna en todas las tareas fue alta (copia sin significado= 0,9676, copia con significado= 0,8750, ensamblaje sin significado= 0,9869, y ensamblaje con significado= 0,9880). Por lo que se concluyó que el test funciona correctamente, de acuerdo con su función diagnóstica, ya que todos los ítems se localizan en bandas superiores de dificultad y discriminación.

INTRODUCCIÓN

La función visoconstruccional es una función viso-espacial[1] que determina, tanto el desempeño gráfico en el dibujo, como el ensamblaje y la construcción de objetos complejos; como la correcta localización de sus componentes (Montañés, Sierra & Matallana, 2011; Simic 2013). Este, a su vez, es un constructo multidimensional que se compone de:

a. Orientación espacial. "la habilidad de percibir relaciones espaciales entre objetos y navegar entre ellos" (Kreutzer, Caplan & DeLuca, 2011, p. 2200).
b. Actividad perceptual. "proceso cortical que resulta en una experiencia sensorial subjetiva, potencialmente significativa o reconocible" (Kreutzer et al., 2011, p.1901).
c. Capacidad motriz. Esta incluye la programación motora y la coordinación.
d. Funciones ejecutivas. "habilidades de nivel superior en las áreas de: planeación, resolución de problemas, atención, flexibilidad mental, iniciación, juicio, inhibición y razonamiento abstracto" (Kreutzer et al., 2011, p. 991).

El déficit en las funciones viso-construccionales se conoce como apraxia construccional. Esta es una alteración en la conexión entre las funciones viso-espaciales y los mecanismos de la actividad motora manual (Trojano & Conson, 2008), y puede aparecer en algunos síndromes, como en la demencia tipo Alzheimer, la demencia por cuerpos de Lewy y la atrofia cortical posterior. Estudios de lesiones cerebrales han confirmado el papel de distintas regiones en las funciones visoconstruccionales; por ejemplo, cuando existe una lesión en las áreas parieto-occipitales, las habilidades visuales se ven afectadas; mientras que cuando se presenta una lesión en las áreas frontales, se ven afectadas las habilidades de programación y comportamiento secuencial (Montañés et al., 2011).

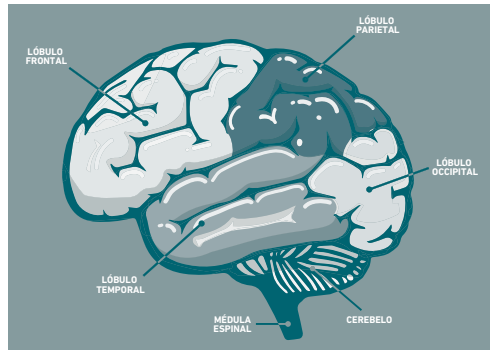


Imagen 2-1 *. Diagrama de caja Pre y Post.*
Fuente: *Elaboración de los autores*

[1] La función visoespacial es el componente de la percepción visual que permite el procesamiento de la orientación y la localización visual de los objetos en el espacio (Kreutzer, Caplan & DeLuca, 2011).

De acuerdo con lo anterior, la prueba fue construida con el objetivo de medir funciones visoconstruccionales; por lo que incluye tareas de copia y ensamblaje con significado y sin significado, diseñadas de acuerdo con la categorización de dos procedimientos propuestos por Grossi y Angelini en 1991 (citados en Montañés et al., 2011): copia con significado (la ruta léxica activa esquemas construccionales conocidos) y copia sin significado (la ruta léxica activa esquemas línea por línea), donde se hace un análisis espacial sin la utilización de esquemas ni representaciones construccionales. Además, en el área clínica, se ha visto que existe un funcionamiento diferencial para objetos vivos e inertes, por lo que dentro de las tareas con significado se incluyeron ambos (Caplan, 2006).

MÉTODO

Muestra

La muestra estuvo conformada por 65 personas (34 hombres, 31 mujeres) entre 16 y 100 años (M=38,37, SD=23,89), sin ninguna condición clínica relevante que pudiera afectar los procesos cognitivos. El objetivo de esta decisión fue la necesidad de estandarizar los puntajes obtenidos de manera que, al ser una prueba diagnóstica, los puntajes bajos correspondieran a cierto grado de deterioro en las funciones visuoconstruccionales. Al mismo tiempo, se hizo muestreo por conveniencia, donde los sujetos se abordaron en las inmediaciones de la Facultad de Ciencias Humanas de la Universidad Nacional de Colombia y en la Fundación Juan Pablo Segundo (personas de la tercera edad).

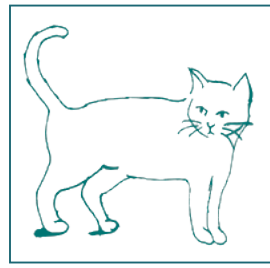
Instrumentos

Test de Orión. Este test consta de dos tipos de tareas (copia y ensamblaje), y dos subtareas (con significado, sin significado) para cada tarea; para este caso en particular, se estructuró como se observa en la tabla 2-1. Igualmente, se solicitó consentimiento informado por parte de cada sujeto.

TAREA	SUBTAREA	IMAGEN UTILIZADA
Copia	Con significado	Gato (imagen 1-1) Silla (imagen 1-2)
	Sin significado	Figuras geométricas (imagen 1-3)
Ensamblaje	Con significado	Rompecabezas de cebra, con su respectivo modelo (imagen 1-4) Rompecabezas de carro, con su respectivo modelo (imagen 1-5)
	Sin significado	Diez palos de balso de diferentes tamaños (entre 1 y 5 cm de largo, por 1 cm de ancho y 1 cm de profundidad) y su respectivo modelo(imagen 1-6)

Tabla 2-1 *Test de Orión, adaptación para el pilotaje.*
Fuente: *Elaboración de los autores*

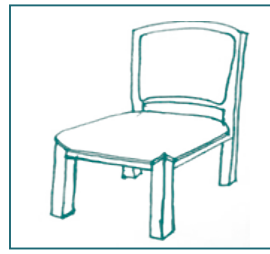
Los ítems de cada una de las tareas se presentan a continuación.



Ítems evaluados

1. Cola
2. Patas traseras
3. Patas delanteras
4. Torso
5. Orejas
6. Ojos
7. Bigotes, nariz y boca
8. Contorno de la cara

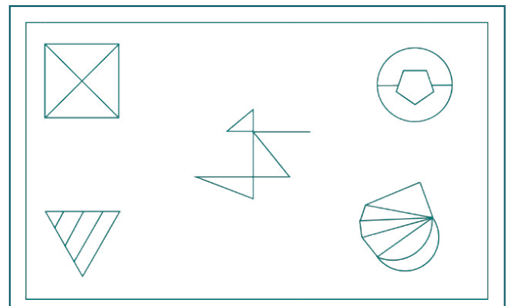
Imagen 2-2. *Ejercicio de copia con significado 1 - Gato.*
Fuente: *Elaboración de los autores*



Ítems evaluados

9. Patas
10. Cojín
11. Espaldar

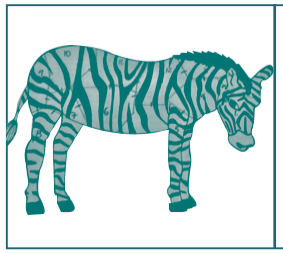
Imagen 2-3. *Ejercicio de copia con significado 2 - Silla.*
Fuente: *Elaboración de los autores*



Ítems evaluados

12. Cuadrado superior izquierdo
13. Líneas transversales al cuadro
14. Triángulo inferior izquierdo invertido
15. Tres líneas dentro del triángulo,
16. Línea horizontal en la figura central
17. Tres triángulos en figura central
18. Círculo superior derecho
19. Pentágono invertido en el centro del círculo
20. Línea horizontal en medio del círculo
21. Tres triángulos pequeños en la figura inferior derecha
22. Triángulo grande en la parte superior de la figura inferior derecha
23. Dos semicírculos en la figura inferior derecha

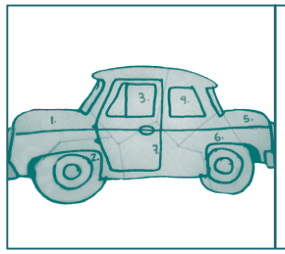
Imagen 2-4. *Ejercicio de copia sin significado – Figuras geométricas.*
Fuente: *Elaboración de los autores*



Ítems evaluados

24. Cabeza
25. Oreja
26. Crin
27. Cuello
28. Patas delanteras
29. Panza
30. Pata trasera izquierda
31. Pata trasera derecha
32. Cola
33. Trasero
34. Lomo trasero
35. Lomo delantero

Imagen 2-5. *Ejercicio de ensamblaje con significado 1 - Cebra.*
Fuente: *Elaboración de los autores*



Ítems evaluados

36. Capó
37. Llanta delantera
38. Ventana delantera
39. Ventana trasera
40. Baúl
41. Llanta trasera
42. Puerta

Imagen 2-6. *Ejercicio de ensamblaje con significado 1 - Carro.*
Fuente: *Elaboración de los autores*

Ítems evaluados

43.Pieza 1
44.Pieza 2
45.Pieza 3
46.Pieza 4
47.Pieza 5
48.Pieza 6
49.Pieza 7
50.Pieza 8
51.Pieza 9
52.Pieza 10

Imagen 2-6. *Ejercicio de ensamblaje sin significado – Palos de balso.*
Fuente: *Elaboración de los autores*

Para realizar validez convergente con las tareas de copia, se utilizó el test de la Figura Compleja de Rey, que consiste en la presentación de una figura de estructura compleja y sin significado, la cual debe ser copiada por el evaluado, teniendo en cuenta la proporción de los detalles y su ubicación. Esta Figura es conocida por evaluar habilidades visoconstruccionales, visoespaciales y de memoria no verbal (Jamus & Mäder, 2005). En neuropsicología clínica, es especialmente utilizada como función diagnóstica.

Procedimiento

Para la construcción de la prueba, se inició con la búsqueda bibliográfica, a fin de generar una red nomológica que permitiría definir el constructo de funciones visoconstruccionales. Posteriormente, teniendo en cuenta la revisión que se realizó, se construyeron las tareas de la prueba; por lo que fue necesario realizar tres búsquedas diferentes: en la primera, enfocada en las tareas de copia, se buscaron imágenes con y significado, que no presentaran muchos detalles para que la tarea fuera de dificultad baja; en la segunda, enfocada en las tareas de ensamblaje con significado, se buscaron imágenes de objetos y animales conocidos; en la tercera, enfocada en las tareas de ensamblaje sin significado, se buscó crear elementos abstractos, los cuales no se asociaran a objetos conocidos. Por lo que, se definieron cuatro tareas para el Test de Orión: Tarea de copia con significado, tarea de copia sin significado, tarea de ensamblaje con significado, y tarea de ensamblaje sin significado. Este test, junto con la Figura Compleja de Rey (FCR) (Rey, 2003), fue aplicado en orden aleatorio, como técnica de contrabalanceo. A manera de regla general, se aplicó la prueba a máximo 6 personas en el mismo momento y lugar.

En cuanto a la calificación, esta re realizó de la siguiente forma:

a. Pruebas de copia. Las pruebas de copia se calificaron entre 0 y 2 puntos, en donde 0 indica que la copia de la sección no cumple con buena ubicación ni es reconocible; 0,5 cuando la copia presenta mala ubicación y mala precisión, pero es reconocible; 1 cuando la copia está bien ubicado, pero tiene mala precisión, y 2 cuando la copia presenta buena ubicación y precisión.

b. En la calificación de las Pruebas de ensamblaje. Se usó un método dicótomo de calificación, en donde 0 indica que la persona ubicó la ficha incorrectamente, y 1 que la persona ubicó la ficha correctamente.

Con el fin de proporcionar evidencias de validez, se realizó la validación por juicio de expertos; la validez convergente (aplicando la FCR a quienes realizaron el Test de Orión), y el análisis de la prueba desde la Teoría Clásica de los Test (TCT). Con respecto a la confiabilidad, se utilizó Alfa de Cronbach para analizar la consistencia interna de la prueba. Finalmente, se utilizaron dos programas para analizar los datos: Lertap, para hacer el análisis de los ítems, y SPSS (versión 22) para correlacionar el puntaje de la prueba en copia sin significado con la prueba Figura Compleja de Rey y estandarizar los puntajes.

RESULTADOS

Evidencia de validez de contenido

Juicio de expertos

Con el fin de proporcionar evidencias de validez por juicio de expertos se utilizó una plantilla propuesta por Escobar Pérez y Cuervo Martínez (2008), para estudiar la validez de contenido, en términos de suficiencia, claridad, coherencia y relevancia. La juez, una estudiante de maestría en psicología aplicada, con énfasis en neuropsicología de la Universidad Nacional, calificó con 4 todos los ítems; esto quiere decir que los ítems cumplen los criterios con un alto nivel.

Análisis de ítems

El análisis de ítems se realizó desde la TCT, por medio del programa Lertap. Puntualmente, se analizó la dificultad teniendo en cuenta la proporción de dificultad de la clave, y la discriminación por comparación de grupos, dividiendo la muestra en 3 grupos, teniendo en cuenta su desempeño (Grupo 1 n=21, grupo 2 n=23, grupo 3 n=21). Los descriptivos de los tres grupos Se muestran en la siguiente tabla:

Grupo	X ítems resueltos	Desviación Estándar	Puntaje Máx.	Puntaje Mín.
1	51,5	0,5	52	51
2	49,9	1,2	51	47
3	27,9	18,5	47	0

Tabla 2-2 *Descriptivos de los grupos por desempeño.*
Fuente: *Elaboración de los autores*

A pesar de la dificultad vista en el conjunto de los ítems del Test de Orión, es de resaltar que la gran mayoría tiene un valor entre 0,8 y 0,9; además, en su totalidad, poseen un valor p [2]>0,60. Esto implica que, en general, los ítems diseñados para la prueba tienen una dificultad alta (desde la TCT), es decir, fueron muy fáciles, y fueron resueltos con una puntuación máxima por muchos de los sujetos evaluados en el pilotaje. Asimismo, los más fáciles pertenecen, en gran medida, a la tarea de Ensamblaje con significado.

Respecto a la discriminación de los ítems, se encontró que todos puntúan entre 0,45 y 0,95. Ello significa que los ítems en general tienen una alta discriminación y los más difíciles están siendo puntuados correctamente para las personas con puntajes totales más altos, en tanto que para aquellas con puntajes totales más bajos suelen ser más puntuados los ítems más fáciles. Se puede ver además, comparando las tablas A y B, que coinciden algunos de los ítems que menos discriminan con aquellos que demuestran ser más fáciles, como el 4 y el 10 (torso del gato y cojín de la silla).

A continuación se hará un breve análisis por tarea de los ítems correspondientes y los puntajes obtenidos:

Copia con significado

Los descriptivos para estas tareas fueron los siguientes:

[2] La 'p' hace referencia al nivel de dificultad que representa la proporción de personas que contestaron correctamente del ítem. Cuando el nivel de dificultad igual o cercano a 1, el ítem es muy fácil de responder; cuando este nivel es igual o cercano a 0, es difícil de responder.

Media	Mediana	Desviación Estándar	Alfa	Error estándar
8,54	10	2,96	0,875	2,1

Tabla 2-3 *Descriptivos de tareas de copia con significado.*
Fuente: *Elaboración de los autores*

Los puntajes más altos y bajos encontrados corresponden a los puntajes más altos y bajos posibles (0-22). El ítem con menor dificultad es el ítem 4 (p=0,62, torso del gato); mientras que el ítem con mayor dificultad es el 1 (p=0,94, cola del gato). Lo anterior se interpreta desde la TCT, en donde se toma un nivel alto de dificultad como la mayor cantidad de éxito (Herrera, Sánchez & Jiménez, 2001). El ítem que más discrimina en esta tarea, en relación con todos los ítems, es el ítem 5 (0,72, orejas del gato), y el que menos lo hace es el 10 (0,45, cojín de la silla) (Ver imágenes 2-1 y 2-2).

Al realizar la comparación entre los grupos de desempeño de la prueba, se observó, en cuanto a la dificultad en general, que los ítems fueron respondidos correctamente por la muestra, siendo el ítem con menor dificultad el 9 (U-Ldiff=0,60), y el ítem con mayor dificultad el 1 (U-Ldiff=0,90). En cuanto a la discriminación, el ítem que menos discriminó fue el 1 (U-Ldisc=0,19) y el ítem que más discriminó fue el 4 (U-Ldisc=0,71).

Copia sin significado

Los descriptivos para esta tarea son:

Media	Mediana	Desviación Estándar	Alfa
9,82	12	3,96	1,42

Tabla 2-4 *Descriptivos de tareas de copia sin significado.*
Fuente: *Elaboración de los autores*

Para esta tarea, los puntajes más altos y bajos encontrados corresponden a los puntajes más altos y bajos posibles (0 y 24). De forma paralela, los ítems con mayor dificultad son el 12, 13, 16 y 18 (p= 0,86) (cuadrado superior izquierdo, líneas diagonales, línea horizontal y círculo superior derecho respectivamente, (ver imagen 2-3); mientras que el ítem con menor dificultad es el 21 (p= 0,72) (triángulo grande). Los ítems que más discriminan en esta tarea son el 14 (p= 0,82, triángulo), el 16 (p= 0,93, línea horizontal) y el 19 (p= 0,86, pentágono).

En el análisis de comparación por grupos de desempeño de la prueba, se observó que en los ítems 13, 14 y 21, para el puntaje 2, los examinados del grupo 2 presentaron un mejor desempeño comparado con el grupo 1 (grupo 1= 0,95, grupo 2= 1 para los ítems 13 y 14, grupo 1= 0,95 y grupo 2= 0,96 para el ítem 21). A pesar de lo anterior, la discriminación de todos los ítems fue positiva, es decir, los evaluados con bajo nivel no están puntuando alto en la

tarea. El ítem con la menor discriminación fue el 13 (U-Ldisc=0,33) y los ítems con mayor discriminación fueron el 21 y el 23 (U-Ldisc=0,71). En cuanto a la dificultad, todos los ítems presentaron un p> 0,5. Esto indica que la mayoría de la muestra total respondió correctamente a los ítems de la tarea. El ítem con la dificultad más baja es el 21 (U-Ldiff= 0,60), mientras que el ítem con la dificultad más alta es el 18 (U-Ldiff=0,81).

Ensamblaje con significado

Los descriptivos para la tarea de ensamblaje con significado son:

Media	Mediana	Desviación Estándar	Alfa
16,38	19	5,91	0,8

Tabla 2-5 *Descriptivos de tareas de copia con significado.*
Fuente: *Elaboración de los autores*

En el análisis se observa que los puntajes más altos y bajos encontrados corresponden a los puntajes más altos y bajos posibles (0-19). El ítem que presenta más dificultad en las tareas de ensamblaje con significado es el número 35 (p=0,80; el cual se denominó como lomo #2) en tanto que los ítems con menor dificultad fueron los 36, 37, 38, 39, 41 y 42, todos con un p=0,89 (correspondientes a la figura del carro). Por otro lado, el ítem más discriminativo es el 40 (0,95) (baúl del carro), y el que menos discrimina para esta tarea, en relación con la totalidad de ítems, es el 31 (p=0,70, pata trasera de la cebra).

En la comparación entre grupos de desempeño de la prueba, se observó que los evaluados con bajo nivel no acertaron en la mayoría de los ítems. Sin embargo, la discriminación no fue alta, siendo la de mayor valor la del ítem 35 (U-Ldisc= 0,57) y la de menor valor la de los ítems 36, 37, 38, 39, 41 y 42 (U-Ldisc=0,3). La dificultad para cada ítem de la tarea fue mayor a 0,5, es decir, que los ítems fueron respondidos correctamente por la mayoría de la muestra total.

Ensamblaje sin significado

Los descriptivos para esta tarea son:

Media	Mediana	Desviación Estándar	Alfa
8,38	10	2,96	0,65

Tabla 2-6 *Descriptivos de tareas de copia sin significado.*
Fuente: *Elaboración de los autores*

Los puntajes más altos y bajos encontrados corresponden a los puntajes más altos y bajos posibles (0 a 10). De forma paralela, en esta tarea, el ítem más difícil es el número 46 (p=0,80), en tanto que los ítems con menor dificultad son el 47, 48 y 50 (p=0,86). Los ítems que más discriminan son el 47, 48 y 50 (p=0,84) (Ver imagen 2-6) y el que menos discrimina es el 52 (p=0,76). El error estándar de la tarea fue el menor comparado con el error de las demás tareas, siendo este de 0,40. Esto indica que el puntaje de los evaluados puede variar 0,40 puntos respecto al puntaje bruto de la tarea.

En la comparación entre grupos, se observó que la discriminación de cada ítem de la tarea fue positiva, en otras palabras, los evaluados con bajo nivel no puntuaron en los ítems. En general, la discriminación de los ítems de la tarea tuvo un nivel aceptable, siendo el ítem con mayor valor el 44 y el 46 (U-Ldisc= 0,57) y la discriminación con menor valor la de los ítems 47, 48 y 50 (U-Ldisc= 0,43). La dificultad para cada ítem de la tarea fue mayor a 0,5, siendo los ítems con menor dificultad el 44 y el 46 (U-Ldiff=0,71 para ambos ítems), y los ítems con mayor dificultad el 47, 48 y 50 (U-Ldiff=0,79) (Ver imagen 2-6).

Evidencia de validez basada en la relación con otras variables

Se correlacionó el puntaje total obtenido por los sujetos en la FCR con el total obtenido por los mismos sujetos en la tarea de copia sin significado, en el Test de Orion; como resultado se obtuvo una correlación significativa positiva, de baja a moderada (r^3=,304, p=,027).

Confiabilidad- Alfa de Cronbach

Se analizó la consistencia interna por medio del alfa de Cronbach, en donde se obtuvo un índice de 0,986 (copia sin significado= 0,9676, copia con significado= 0,8750, ensamblaje sin significado= 0,9869, y ensamblaje con significado= 0,9880). Los resultados indican que las tareas son fiables y presentan alta consistencia interna, ya que un alto coeficiente de cronbach se relaciona con una poca varianza en los ítems, en relación con la varianza total (Herrera, 1998).

Estandarización

Para estandarizar los resultados, se transformó el puntaje bruto a puntaje Z, y, a partir de los percentiles, se realizaron baremos de acuerdo con la edad de la muestra, formando 3 rangos (1= 16-30, 2= 31-65, 3=66-100).

DISCUSIÓN

A pesar de que se realizó un análisis factorial exploratorio (rotación varimax), debido al tamaño de la muestra y la poca variabilidad de la misma muestra, esta no se tuvo en cuenta como evidencia de validez de estructura interna de la prueba.

Con respecto al análisis de ítems, el hecho de que en 24 de los ítems ninguna persona obtuviera un puntaje de 0,5 o 0 en las tareas indica que no

[3] La 'r' hace referencia al índice de correlación de Pearson.

se encontraron personas con niveles bajos de aptitud, lo que generó que la dificultad fuera muy alta, al igual que la discriminación. Esto se dio ya que el pilotaje se aplicó exclusivamente en población en óptimas condiciones de salud, y, al ser una prueba de diagnóstico utilizada en población clínica, la población normal debería puntuar exclusivamente en valores superiores en la prueba para que esta tenga capacidad de discriminación. Se explica, entonces, que este se presentó debido a un error en el pilotaje, donde debió incluirse población clínica, la cual sería la que teóricamente tendría la tendencia a tener puntajes bajos. Las razones anteriores también explican la escasa puntuación en los puntajes 0 y 0,5, así como la alta varianza en los puntajes del grupo 3.

Los ítems de toda la prueba están discriminando positivamente, es decir, las personas que tienen un bajo nivel de habilidades visoconstruccionales no obtuvieron una puntuación alta en la prueba. Así como se observó la dificultad por grupos para todas las tareas, con respecto al promedio de las opciones, se evidenció que ningún ítem de los examinados con puntajes bajos globales lograron puntales altos (puntaje de 2 o 1 en cada ítem, según sea copia o ensamblaje) en las tareas.

Con relación a esto último, se encontró que las personas en el grupo con bajo nivel obtuvieron puntaje de 1, con mayor frecuencia para la copia. Esto se debe a que el puntaje 2 indica que la tarea fue realizada correctamente, mientras que el puntaje 1 indica que la tarea, aunque se logró realizar, no cumplió con el requisito de precisión o ubicación de lo copiado. Empero lo anterior, para algunos ítems de la tarea de dibujo sin significado (13, 14 y 21 según la prueba total), las personas del grupo 2 puntuaron más que las personas del grupo 1, este fenómeno se debe a un dato extremo que afectó el promedio.

Con respecto a la correlación con la Figura Compleja de Rey, se encontró que, aunque significativa, la correlación fue de baja a moderada. Esto puede deberse a que la FCR solo mide copia sin significado, mientras que el Test de Orión está midiendo tanto copia con y sin significado, como ensamblaje con y sin significado. Se puede decir que la operacionalización del constructo es distinta para ambas pruebas, y por esto la correlación entre los puntajes totales no es tan alta, aunque sí significativa.

Según lo calificado por la juez, todos los ítems son suficientes para obtener la medición de las dimensiones; son claros, es decir, la instrucción se comprende fácilmente ya que tiene buena redacción y semántica; son coherentes con la dimensión que se está midiendo; y son todos relevantes, el eliminarlos de la prueba afectaría la medición de la dimensión.

DECISIONES Y LIMITACIONES

En general, el Test de Orión muestra un buen funcionamiento como prueba para identificar personas sin algún tipo de lesión o afectación cognitiva, ya que todos sus ítems se localizan en rangos de dificultad superiores, siendo fáciles para la población normal; no obstante, todos sus ítems se localizan en rangos de discriminación superior, por lo que se puede discriminar entre grupos saludables y clínicos. Además, la prueba muestra adecuada consistencia interna y evidencia de validez de la estructura interna, estando los factores apoyados en la teoría.

La prueba presentó algunas limitaciones; especialmente, los materiales empleados para la misma. Por ende, se recomienda hacer las fichas del rompecabezas de un material diferente a cartón paja o cartulina, ya que muchas veces los participantes se demoraban más tiempo en terminar la figura y en algunos casos se rendían debido a que era difícil manipular las fichas, esto especialmente en la cebra.

Con respecto a la toma de decisiones sobre los ítems, al encontrarse en el análisis de comparación entre grupos por desempeño, ítems con dificultad muy alta, y con muy baja discriminación (1, 6, 8, 13, 36 y 37), se podría pensar que estos ítems no están funcionando dentro de la prueba, por lo que se debería hacer análisis de funcionamiento diferencial o eliminar los ítems de la prueba, si se mantiene la misma discriminación cuando se realice el pilotaje con muestra clínica. Para futuros estudios, se recomienda ampliar la muestra para poder realizar análisis factorial.

Debido a que la prueba piloto estuvo limitada a población saludable, el grupo de menor desempeño estuvo sobrepresentado al no aplicarse la prueba a población clínica, lo cual estaría afectando la dificultad y la discriminación. Entonces, no se puede asegurar que los ítems estén funcionando o no con base en la comparación entre grupos por desempeño, si los grupos no están siendo representativos.

Como opción, con respecto a la calificación, se recomienda incluir el tiempo en el puntaje de las tareas ya que esta variable brinda mayor información respecto a las diferencias entre grupos. A su vez, se propone utilizar un puntaje derivado del tiempo total necesario para realizar la prueba y no solo basarse en el acierto o falla. También se podría desarrollar una medición parcial del tiempo, es decir, que cada cierto tiempo se anote el rendimiento de la persona en la prueba.

REFERENCIAS

Caplan, L. (2006). Art, constructional apraxia, and the brain. International Review of Neurobiology. The Neurobiology of Painting, (74), 215-232. DOI: 10.1016/S0074-7742(06)74015-6.

Escobar-Pérez, J. & Cuervo-Martínez, A. (2008). Validez de contenido y juicio de expertos: una aproximación a su utilización. Avances en Medición. 6(1), 27-36.

Herrera, A. N. (1998). Notas sobre psicometría. Santafe de Bogotá: Universidad Nacional de Colombia. Recuperado de: https://bit.ly/2s7Os2F

Herrera, A. N., Sánchez, N. & Jiménez, H. (2001). De la teoría clásica de los test a la teoría de respuesta a los ítems. En: J. I. Ruiz Pérez, E. Ponce de León Díaz, A. N. Herrera, N. Sánchez, H. Jiménez & E. Medellín Lozano (Eds.). Avances en Medición y evaluación en Psicología y Educación: cinco lecturas selectas. Aula Psicológica (pp. 293-332). Bogotá: Universidad El Bosque.

Jamus, D. R. & Mäder, M. J. (2005). A figura complexa de Rey e seu papel na avaliação neuropsicológica. Jornal of Epilepsy and Clinical Neurophysiology. 11(4). DOI: 10.1590/S1676-26492005000400008

Kreutzer, J. S., Caplan, B., & DeLuca, J. (Eds.). (2011). Encyclopedia of clinical neuropsychology. New York: Springer.

Montañés, P., Sierra, N., & Matallana, D. (2011). Habilidades viso-construccionales y modelos de la neuropsicología cognoscitiva. En Montañés, P. Demencias y la neuropsicología del dibujo. (pp. 89-122). Bogotá, Colombia: Pontificia Universidad Javeriana.

Rey, A. (2003). Manual Test de Copia y de Reproducción de Memoria de Figuras Geométricas Complejas (8a Edición), TEA Ediciones

Simic, N., Khan, S., Rovet, J. (2013). Visuospatial, Visuoperceptual and Visuoconstructive Abilities in Congenital Hypothyroidism. Journal of the Internacional Neuropsychological Society, 19, 1119 – 1127.

Trojano, L. & Conson, M. (2008). Visuospatial and visuoconstructive deficits. Handbook of Clinical Neurology (Vol. 88, pp. 373-391). DOI: 10.1016/S0072-9752(07)88019-5.

ANEXO 1

Planilla Juicio de Expertos

Respetado juez,

Usted ha sido seleccionado para evaluar el instrumento Test de Orión, el cual hace parte del proyecto final de la materia de Psicometría, dictada por la profesora Aura Nidia. La evaluación de los instrumentos es de gran relevancia para lograr que sean válidos y que los resultados obtenidos a partir de estos sean utilizados eficientemente; aportando tanto al área investigativa de la Psicología como a sus aplicaciones. Agradecemos su valiosa colaboración.

Nombres y apellidos del juez: Liliana Duarte Pedroza
Formación académica: Psicóloga, estudiante Maestría en Psicología (4 semestre)
Áreas de experiencia profesional: Neuropsicología
Tiempo: 2 años **Cargo actual:** Psicóloga
Institución: Servicio de Evaluación Neuropsicológica, Universidad Nacional de Colombia
Objetivo de la investigación: Validación del instrumento.
Objetivo del juicio de expertos: Demostrar el tipo de evidencia de validez del contenido de forma cualitativa para el instrumento.
Objetivo de la prueba: Medir las funciones visoconstruccionales con fines diagnósticos.

CATEGORÍA	CALIFICACIÓN	INDICADOR
SUFICIENCIA Los ítems que pertenecen a una misma dimensión bastan para obtener la medición de esta.	1. No cumple con el criterio 2. Bajo criterio 3. Moderado nivel 4. Alto nivel	a. Los ítems no son suficientes para medir la dimensión. b. Los ítems miden algún aspecto de la dimensión pero no corresponden con la dimensión total. c. Se deben incrementar algunos ítems para poder evaluar la dimensión completamente. d. Los ítems son suficientes.
CLARIDAD La instrucción se comprende fácilmente, es decir, su sintáctica y semántica son adecuadas.	1. No cumple con el criterio 2. Bajo criterio 3. Moderado nivel 4. Alto nivel	a. Las instrucciones no son claras. b. Las instrucciones requieren bastantes modificaciones o una modificación muy grande en el uso de las palabras de acuerdo con su significado o por la ordenación de las mismas. c. Se requiere una modificación muy específica de algunos de los términos de las instrucciones. d. Las instrucciones son claras, tienen semántica y sintaxis adecuada.

COHERENCIA El ítem tiene relación lógica con la dimensión o indicador que está midiendo.	1. No cumple con el criterio 2. Bajo criterio 3. Moderado nivel 4. Alto nivel	a. El ítem no tiene relación lógica con la dimensión. b. El ítem tiene una relación tangencial con la dimensión. c. El ítem tiene relación moderada con la dimensión que está midiendo. d. El ítem se encuentra completamente relacionado con la dimensión que está midiendo.
RELEVANCIA	1. No cumple con el criterio 2. Bajo criterio 3. Moderado nivel 4. Alto nivel	a. El ítem puede ser eliminado sin que se vea afectada la medición de la dimensión. b. El ítem tiene alguna relevancia, pero otro ítem puede estar incluyendo lo que mide este. c. El ítem es relevantemente importante. d. El ítem es muy relevante y debe ser incluido.

ANEXO 2

Calificación juicio de expertos

Tarea	Subtarea	Item		Suficiencia	Coherencia	Relevancia	Claridad	Observaciones
COPIA	CON SIGNIFICADO	GATO	1	4	4	4	4	Ninguna
			2	4	4	4	4	
			3	4	4	4	4	
			4	4	4	4	4	
			5	4	4	4	4	
			6	4	4	4	4	
			7	4	4	4	4	
			8	4	4	4	4	
		SILLA	9	4	4	4	4	Ninguna
			10	4	4	4	4	
			11	4	4	4	4	
	SIN SIGNIFICADO	FIGURAS GEOMÉTRICAS	12	4	4	4	4	La figura inferior derecha podría tener una menor complejidad visual, ya que, en comparación con las otras 3, su complejidad es mayor y esto podría tener algún impacto sobre el desempeño.
			13	4	4	4	4	
			14	4	4	4	4	
			15	4	4	4	4	
			16	4	4	4	4	
			17	4	4	4	4	
			18	4	4	4	4	
			19	4	4	4	4	
			20	4	4	4	4	
			21	4	4	4	4	
			22	4	4	4	4	
			23	4	4	4	4	
ENSAMBLAJE	CON SIGNIFICADO	CEBRA	24	4	4	4	4	Ninguna
			25	4	4	4	4	
			26	4	4	4	4	
			27	4	4	4	4	
			28	4	4	4	4	
			29	4	4	4	4	
			30	4	4	4	4	
			31	4	4	4	4	
			32	4	4	4	4	
			33	4	4	4	4	
			34	4	4	4	4	
			35	4	4	4	4	
		CARRO	36	4	4	4	4	Ninguna
			37	4	4	4	4	
			38	4	4	4	4	
			39	4	4	4	4	
			40	4	4	4	4	
			41	4	4	4	4	
			42	4	4	4	4	
	SIN SIGNIFICADO	PALOS DE BALSO	43	4	4	4	4	Ninguna
			44	4	4	4	4	
			45	4	4	4	4	
			46	4	4	4	4	
			47	4	4	4	4	
			48	4	4	4	4	
			49	4	4	4	4	
			50	4	4	4	4	
			51	4	4	4	4	
			52	4	4	4	4	

¿Hay alguna dimensión que hace parte del constructo y no fue evaluado?

SÍ: ___ No: **X**

¿Cuál?: NA

** Para los casos de equivalencia semántica se deja una casilla por ítem, ya que se evaluará si la traducción o el cambio de vocabulario son suficientes.*

Observaciones finales:

Tareas:

La prueba en conjunto logra abordar, de manera adecuada, las distintas dimensiones requeridas para una correcta evaluación de las habilidades práxicas visoconstruccionales. La prueba podría tener utilidad clínica.