

TEST RAVEN2. ANÁLISIS PRELIMINARES PARA LA CONSTRUCCIÓN DEL BAREMO DE ADULTOS ENTRE 19 Y 64 AÑOS DE LAS PROVINCIAS DEL NORDESTE ARGENTINO

Ana Kohan Cortada¹, Daniela Elorrieta², Gabriela Spada², Cecilia Martínez², Dorita Ceberio Macías² y Rosiel Pardo²

RESUMEN

El objetivo de este trabajo es presentar los análisis preliminares que servirán como base para la elaboración de un baremo del Test Raven2, para adultos de entre 19 y 64 años residentes en las provincias del Nordeste Argentino (NEA). Este proyecto cuenta con el respaldo financiero de la Universidad de la Cuenca del Plata (UCP) y está desarrollado por un equipo interdisciplinario compuesto por docentes, gestores, becarios y estudiantes de Psicología, comprometidos con el diseño de pruebas psicológicas y su aplicación ética y confiable, bajo la dirección de la Dra. Ana Kohan Cortada.

El Raven es un test altamente utilizado en Argentina, sin embargo, se ha constatado que no hay baremos en la región del NEA (Kohan Cortada et al., 2024). Por la gran difusión que han tenido los ítems de las distintas versiones utilizadas del Raven a lo largo del tiempo, se considera útil estudiar esta nueva versión que posee todos sus ítem nuevos y generados con la Teoría de Respuesta al Ítem (TRI) para evaluar las capacidades cognitivas. Se necesita contar con herramientas estandarizadas, válidas, confiables y que las extranjeras a utilizar posean baremos locales, como prueba eficaz para la evaluación de la capacidad intelectual y su aplicación en los ámbitos de la psicología clínica, educacional, jurídica y laboral.

El equipo administró el Raven2 a 917 personas de un total de 1680 participantes previstos de diferentes localidades del NEA. La muestra estuvo integrada por un 68% de mujeres y el 32% hombres, con una edad promedio de 31.30 ± 11.23 , de distintos niveles educativos, todos participantes de la población general que dieron su consentimiento informado.

Se llevaron a cabo análisis de la consistencia interna del test, el índice de dificultad, el índice de discriminación y la correlación Biserial–Puntual para cada ítem del test.

Palabras clave: Adultos, Confiabilidad, Nordeste argentino, Raven2.

¹ Centro interdisciplinario de investigaciones en Psicología Matemáticas y Experimental (CIIPME–CONICET). Instituto de Investigaciones Científicas (IDIC). Universidad de la Cuenca del Plata (UCP). Salguero 1692 8° CABA anakohancortada@yahoo.com.ar. anakohancortada@conicet.gov.ar.

² Instituto de Investigaciones Científicas (IDIC). Universidad de la Cuenca del Plata (UCP).

ABSTRACT

The objective of this work is to present the preliminary analyses that will serve as a basis for the elaboration of a scale of the Raven2 Test for adults between 19 and 64 years of age living in the provinces of Northeastern Northeast Argentina (NEA). This project is financially supported by the Universidad de la Cuenca del Plata (UCP) and is developed by an interdisciplinary team composed of teachers, managers, scholarship holders and Psychology students, committed to the design of psychological tests and their ethical and reliable application, under the direction of Dr. Ana Kohan Cortada.

The Raven is a test highly used in Argentina, however, it has been found that there are no scales in the NEA region (Kohan Cortada et al., 2024). Due to the great diffusion that the items of the different versions of the Raven have had over time, it is considered useful to study this new version, which has all its items new and generated with the Item Response Theory (IRT) to assess cognitive abilities. It is necessary to have standardized, valid, reliable tools and that the foreign ones to be used have local scales, as an effective test for the evaluation of intellectual capacity and its application in the fields of clinical, educational, legal and work psychology.

The team administered the Raven2 to 917 people out of a total of 1680 planned participants from different locations in the NEA. The sample is composed of 68% women and 32% men, with an average age of 31.30 ± 11.23 , from different educational levels, all participants from the general population who have given their informed consent.

Analyses of the internal consistency of the test, the difficulty index, the discrimination index and the Biserial–Point correlation were carried out for each test item.

Keywords: *Adults, Reliability, Northeast Argentina, Raven2.*

INTRODUCCIÓN

En la evaluación psicológica es necesario emplear herramientas estandarizadas, válidas y confiables. También, a la hora de utilizar instrumentos extranjeros, los mismos se deben adaptar, validar, así como construir baremos locales para la toma de decisiones en los distintos ámbitos: clínico, educativo, laboral o jurídico (International Test Commission [ITC], 2000, 2017).

En 2018, se terminó de construir en Estados Unidos el Raven2. Esta nueva versión presenta todos sus ítems nuevos pero se conservan los constructos a medir.

El objetivo de este artículo es comunicar los análisis preliminares realizados que permitirán avanzar hacia la construcción del baremo. En este trabajo se analizaron las propiedades psicométricas del test Raven2 con una muestra no probabilística de 917 adultos de entre 19 y 64 años. Se llevaron a cabo análisis de la consistencia interna del test, así como también la proporción de respuesta correcta por distractor, el índice de dificultad, el índice de discriminación y la correlación Biserial-Puntual para cada ítem del test.

El proyecto está financiado completamente por la Universidad de la Cuenca del Plata. El equipo de trabajo está integrado por docentes, gestores, becarios y alumnos de la carrera de psicología y es dirigido por la Dra. Ana Kohan Cortada del CIIPME-CONICET.

El Test de Raven es una de las pruebas más utilizadas mundialmente. En 1938 J. Raven construye el Test de Matrices Progresivas para reflejar el factor G de Spearman o inteligencia general, que estaba constituida por dos componentes identificados como habilidad educativa y habilidad reproductiva (Raven et al., 1993), afirmando que su prueba específicamente mide la capacidad educativa.

El test propone problemas de dificultad y complejidad creciente. Se trata de una prueba no verbal ya que no requiere respues-

tas orales ni escritas, además utiliza formas geométricas universales, que no dependen de la formación ni de la cultura para identificarse. Esto permite que la prueba pueda ser realizada independientemente del nivel cultural y de conocimientos que posea quien la ejecuta.

Hasta 2018 existían tres versiones del test: Las escalas General, Coloreada y Avanzada. Todas requieren que el sujeto ponga en marcha el proceso de la educación, implica que la persona realice inferencias de las relaciones entre reactivos abstractos, que pueda dar sentido al desorden, buscar lo que no es obvio y formarse ideas que ayuden a afrontar problemas complejos. Esta capacidad se relaciona con diversas funciones cognitivas como la atención visual, aptitud de clasificación, razonamiento inductivo, memoria operativa, aptitud espacial (Pearson Inc, 2019).

El nuevo Raven2:

Esta nueva versión del test posee todos sus ítems nuevos (tanto para los formatos online y como papel), se han calibrado en base a la TRI, son en color, con un total de 5 series (A, B, C, D, y E), cada serie posee 12 problemas a resolver. Ofrece 5 opciones para dar la respuesta a cada problema, por lo tanto, el cálculo de la probabilidad de responder por azar es la misma a lo largo de toda la prueba.

En esta versión las 3 primeras series (A, B y C) sirven para administrar a niños de 4 a 8.11 años de edad, mientras que para los individuos de 9 a 69 años se usan las series B a la E.

El formato online se puede aplicar a través de la plataforma de Pearson Clinical & Talent Assessment (Q-global), lo que agiliza tanto el tiempo de corrección como la obtención del informe de cada sujeto, que es inmediata. Este formato posibilita administrarlo en dos versiones según la cantidad de ítems, una forma larga y otra corta (48 y

24 ítems, respectivamente). Una vez que se elige la cantidad de ítems a aplicar el sistema asigna de forma automática un grupo único de ítems aleatoriamente extraídos de un banco de 329 ítems. Dichos ítems fueron generados sobre la base de la Teoría de Respuesta al Ítem (Lord, 1980; Rasch, 1963).

La Teoría de Respuesta al Ítem (TRI) se centra en entender cómo las respuestas de una persona a una pregunta de un test se relacionan con un rasgo que se está midiendo, como puede ser el nivel de conocimiento o habilidad de la persona. Para esto, se utilizan fórmulas matemáticas que calculan la probabilidad de que alguien responda correctamente según su nivel de habilidad.

El objetivo principal de la TRI es crear herramientas de medición que funcionen igual de bien para diferentes grupos de personas. Esto significa que dos personas con el mismo nivel de habilidad deberían tener la misma probabilidad de responder correctamente una pregunta, sin importar de qué grupo social, cultural o educativo provengan.

A diferencia de la Teoría Clásica de los Test (TCT), que se basa en el puntaje total de la prueba, la TRI analiza cada pregunta de forma individual y se enfoca en medir el rasgo que se está evaluando. Esto permite que, al medir el nivel de habilidad de una persona a partir de sus respuestas, esa medición sea precisa incluso si cambian las preguntas del test. Es decir, la TRI permite obtener resultados más exactos sin depender de un grupo específico de personas o de las características de las preguntas del test.

La TRI tiene en cuenta tres aspectos clave: la dificultad de las preguntas, la habilidad de la persona y la probabilidad de que una persona con cierto nivel de habilidad responda correctamente o no a una pregunta.

En esta nueva versión del Raven2 tanto los ítems de sus dos formas digitales, corta y larga, como la que se presentan en formato papel son equivalentes (Cortada de Kohan, 2004).

En este trabajo, dado que es el primer estudio realizado con el Raven2 en nuestro país, se ha decidido llevar a cabo los análisis descritos con el propósito de evaluar el rendimiento de los ítems y las características psicométricas globales del test. Esto representa un paso preliminar esencial para la adaptación y baremación del instrumento en el contexto específico del nordeste argentino.

MATERIALES Y MÉTODO

El presente estudio es cuantitativo de tipo instrumental, según la clasificación (Ato et al., 2013), ya que se trata de un estudio tendiente a la obtención de evidencias sobre el funcionamiento de la prueba en nuestro medio, específicamente en el NEA; y psicométrico, como información preliminar para avanzar hacia la obtención de un baremo, como herramienta que permita aplicarse en nuestro medio para la evaluación psicológica, específicamente de la capacidad cognitiva.

Se utilizó el Raven2 en formato papel. Se trabajó con población adulta. Población y muestra:

El universo sobre el cual se obtuvo la muestra está compuesto por sujetos de población general, entre 19 y 64 años, residentes en los distintos conglomerados del NEA (Formosa, Posadas, Corrientes, Resistencia, Sáenz Peña y zonas urbanas y rurales vecinas).

La muestra fue de tipo no probabilística. Al momento la muestra está constituida por 917 sujetos evaluados de los 1680 planificados.

Procedimiento para la recolección de datos:

El procedimiento seguido contempló las normativas profesionales internacionales para la adaptación y la validación de tests usados en la práctica clínica e institucional (ITC, 2000).

Durante el desarrollo del trabajo se respetaron los principios éticos de la investigación con seres humanos, procurándose las condiciones necesarias para proteger el anonimato y la confidencialidad de los datos. Los participantes accedieron voluntariamente a colaborar con la investigación después de haber firmado el consentimiento informado.

Para implementar la aplicación del Raven2 los participantes fueron contactados por los integrantes del equipo de investigación, en el que participan docentes, gestores, alumnos avanzados y becarios de la UCP especialmente capacitados para administrar la técnica. Los mismos fueron entrenados con el fin de garantizar la homogeneidad de los procedimientos.

Se convocó a participantes entre los 19 y 64 años, utilizando redes sociales, flyers, mailing, formularios de Google, se invitó a organizaciones que tuvieran convenios con la Universidad y, se utilizó la recomendación personal de los docentes, becarios, alumnos y gestores que participan en la investigación, así como, los mismos sujetos incluidos en el estudio, que luego de su participación recomendaron la experiencia sumando más voluntarios.

Se elaboró un instructivo estandarizado donde se explicitaba de qué se trataba la investigación y la importancia de su participación.

Aquellos sujetos que aceptaron fueron citados en grupos, que no superaban las 13 personas por vez, en las instalaciones de la Universidad o bien en las instituciones conveniadas, indicándoles que toda la aplicación llevaría unos 90 minutos como máximo para que puedan responder de forma grupal en un espacio que cumplía con las condiciones de iluminación y comodidad destinado a tal fin.

Antes de comenzar con la administración, una vez reunidos junto con el equipo de investigación, se les volvió a explicar a los voluntarios el objetivo del proyecto y se leyeron las indicaciones para la firma del consentimiento informado.

A cada participante se le entregaron

tres hojas con un mismo código impreso en el ángulo superior derecho, lo cual permitió separar a la primera (consentimiento informado y datos de contacto del sujeto) de las otras dos, cuestionario sociodemográfico básico y la hoja de respuestas del Raven2.

Luego, los integrantes del equipo de investigación entregaron el cuadernillo de estímulos junto con un lápiz con goma a cada participante.

La consigna para dar comienzo se dio simultáneamente, los administradores permanecieron en el salón de aplicación atentos, durante cada encuentro, para poder desplazarse y dar respuesta a cualquier duda que pudiera aparecer de los participantes no relacionada con la resolución de los problemas. Se les solicitó a todos los participantes que mantuvieran apagados los celulares para impedir que los sujetos fotografiaran los estímulos.

Una vez finalizada la toma, y que no quedara nadie en el salón, se hacía pasar al siguiente grupo.

Se llevó a cabo la recolección de los datos siguiendo los procedimientos y consignas ya estandarizadas para la administración. En cada sede se contó con un docente responsable del equipo para resguardar, controlar y ordenar los protocolos.

Al finalizar la administración los integrantes del equipo verificaron que estuviera completo el material entregado, y separaron la primera hoja de las restantes para ser guardadas en un sobre cerrado, con lo cual todos los datos personales quedan protegidos.

Criterios de inclusión de los participantes: Audición y visión normal (o corregida), no estar bajo tratamiento psiquiátrico o tomando medicación que pudiera afectar el rendimiento en la prueba, ausencia de enfermedad que pudiera afectar el rendimiento cognitivo durante la prueba, no haber sido diagnosticado con una afección neurológica (cirugías, tumores, encefalitis, etc.), no haber sido evaluado en cuanto a la aptitud cognitiva

con pruebas similares en los últimos 6 meses antes de la fecha de aplicación del Raven2.

Todos los protocolos completos se ordenaron y numeraron, para luego volcar los datos de la información adquirida en una matriz de datos construida especialmente para su posterior análisis estadístico.

Para la carga de datos los miembros del equipo utilizaron en programa Excel de Microsoft y para los análisis estadísticos más complejos se trabajó con el paquete IBM–SPSS 24 o bien el Jamovi.

La descripción de las variables género y nivel educativo (dicotomizado en dos niveles: hasta 12 años y más de 12 años de educación), se llevó a cabo por medio de frecuencias absolutas y relativas simples. En el caso de la variables edad, para su descripción se empleó el cálculo de medidas de tendencia central (Media) y dispersión (desviación estándar); así como el cálculo de las frecuencias absolutas de cada grupo de edad considerado (19 a 24 años, 25 a 34 años, 35 a 44 años, 45 a 54 años, 55 y más) en función del género. Con respecto a la puntuación total obtenida en el Raven2, para la misma se calcularon los descriptivos Media, Desviación estándar (DE), Mediana (Md), valor Máximo (Máx) y Mínimo (Mín) en función del género y nivel educativo dicotomizado (hasta 12 años y con más de 12 años de educación).

Seguidamente, se realizó un análisis preliminar de los datos a modo exploratorio para verificar los supuestos estadísticos y también detectar casos atípicos que pudieran

distorsionar los resultados.

Se realizaron los análisis psicométricos de confiabilidad utilizando la prueba Kuder–Richardson (KR–20).

Luego, se realizó un análisis del desempeño de los ítems individuales del test, considerando la cantidad de respuestas correctas por ítem y por opción de respuesta (índice de dificultad, valor p), también se calculó el índice de discriminación (valor d) y la correlación biserial puntual (ítem–test, R_{bp}) para evaluar el poder discriminativo de los mismos, es decir, su capacidad para distinguir entre quienes evidencian niveles elevados y quienes evidencian niveles bajos en el criterio.

RESULTADOS

A continuación se presentan los resultados preliminares del estudio.

La muestra hasta el momento posee 917 participantes.

En relación con la edad, los 917 participantes informaron tener entre 19 y 64 años (Edad: $M = 31.29$; $DE = 12.20$). La cantidad de mujeres en la muestra fue de 626 (68.04%) y 291 (31.73%) hombres. La variable educación se dicotomizó teniendo en cuenta aquellos que tenían hasta 12 años y más de 12 años de escolaridad, 39,58% y 60,41% respectivamente.

La distribución de género por edad agrupada se puede apreciar en la Tabla 1.

Tabla 1
Género y Edad

Género	EDAD					Total
	19 a 24	25 a 34	35 a 44	45 a 54	55 y más	
Hombre	130	66	48	32	15	291
Mujer	268	149	103	57	49	626
Total	398	215	151	89	64	917

En la Tabla 2 se puede apreciar el educativo dicotomizado.
puntaje Total del Raven2 por género y nivel

Tabla 2

		MEDIA	DE	Md	Máx.	Mín.	N
Género	Mujer	23.68	6.7	24	43	0	626
	Hombre	25.39	7.93	24	45	0	291
N. Educ.	<=12 años	21.08	7.25	21	44	0	363
	> 12 años	25.65	6.77	25	45	0	554

Funcionamiento del instrumento:

Con el objetivo de estudiar el funcionamiento del instrumento se llevó a cabo el estudio de la fiabilidad calculando el índice de Kuder–Richardson–20, que es un tipo especial de alfa de Cronbach para escalas con respuestas dicotómicas (Kuder y Richardson, 1937). El mismo expresa el grado de correlación promedio entre todas las divisiones por mitad posibles de un instrumento de medición, asumiendo que estos ítems miden la misma variable que dicen medir.

Se realizó el análisis de los 48 ítems (series B, C, D y E) y alcanzó un índice KR–20 = .865, lo que evidencia una consistencia interna muy buena.

También para estudiar el funcionamiento de los nuevos ítems del Raven2 en el NEA se calcularon los índices de dificultad y de discriminación.

El índice de dificultad indica la prevalencia de respuestas correctas por ítem, sirve para evaluar la adecuación de las opciones de respuesta y el nivel de dificultad de los ítems.

Dicho índice expresa la proporción de sujetos que respondieron correctamente al ítem sobre el total de sujetos que hicieron el test (p). Los valores que presenta oscilan entre 0 y 1, siendo la dificultad de ítem menor cuanto mayor sea el valor del índice (Tabla 3).

Entre los valores obtenidos, aquellos que oscilaron entre .30 y .70 fueron los óptimos, pues indicaron que no serían ítems ni excesivamente fáciles ni excesivamente difíciles de resolver.

En la Tabla 3 se puede observar, respecto del índice de dificultad, que en la serie B se obtuvieron valores muy altos de p , lo que indica la facilidad de resolución del ítem; pero no sucede lo mismo con la serie E, en donde los ítems resultaron sumamente difíciles para los participantes. Se ha de recordar que esta nueva versión del Raven2 integra ítems de las tres versiones anteriores, es decir contiene ítems con las características de las escalas coloreada, general y avanzada.

En base a los resultados encontrados, y siguiendo a Cortada de Kohan (1999), la clasificación de los ítems del Raven2 por el índice de dificultad se agruparían en:

- Muy fácil (.81 a 1): B1 al B12.
- Relativamente fácil (.66 a .80): C: 1, 2, 3, 5 y D: 1, 5
- Dificultad adecuada (.51 a .65): C: 4, 6, 8 y D: 6, 7
- Relativamente difícil (.31 a .50): C: 7, 10; D: 2, 3, 4, 8, 9, 10, 11 y E: 7
- Difícil (.11 a .30): C: 9, 11, D: 12, E: 1 al 6 y 9

- Muy difícil (0 a .10): C: 12 y E: 8, 10, 11 y 12.

También se llevó a cabo el cálculo del índice de discriminación (d). Dicho índice permite identificar la capacidad del ítem para distinguir entre aquellos que muestran niveles altos y bajos en el criterio (d).

Con respecto a la discriminación de un ítem se han seguido las pautas sugeridas por la literatura actual (Lloret–Segura et al., 2014). Para el cálculo de este índice de discriminación (d) se emplearon pruebas de contraste T de Student (Backhoff et al., 2000; Cortada de Kohan, 1999), mediante los cuales se compararon la cantidad de personas con puntajes altos en el test que respondieron correctamente a cada ítem (grupo superior), con la cantidad de personas con puntajes bajos en el test que respondieron correctamente al mismo ítem (grupo inferior).

Los grupos superior e inferior se establecieron considerando el 27% superior e inferior de la muestra completa respectivamente, a partir del puntaje general (Media=24.57 y DE=7.2). El grupo superior estuvo formado por los participantes con un puntaje mayor o igual de 29, mientras que para el grupo inferior fue un puntaje menor o igual a 19 puntos.

En la Tabla 3 se pueden observar los valores d para cada uno de los ítems. Al emplear los valores T de Student como indicadores de este índice, hay que señalar que cuanto más alto sea el índice de discriminación (d), el ítem diferenciará mejor a las personas con puntajes altos y bajos.

Hay que señalar que todos los ítem resultaron estadísticamente significativos con $p < .05$ a excepción de tres ítem que pertenecen a la Serie E (10, 11 y 12). Si bien los índices de discriminación encontrados con esta muestra oscilaron entre los valores .008 y .669, siguiendo la clasificación propuesta por Díaz Rojas y Leyva Sánchez (2013), se

puede decir que tienen:

- Excelente poder discriminativo ($d > .35$): B: 7, 11, 12, C: 1 a 10, D: 1 a 11 y E: 2, 7.
- Buen poder discriminativo (d entre .26 y .35): B: 4, 6, 8, D: 12 y E: 1, 3, 5, 6.
- Regular poder predictivo (d entre .15 y .25): B: 5, 9, 10, C: 12 y E: 9.
- Pobre poder predictivo (d entre 0 y .14): B: 1 a 3; C: 11 y D: 4, 8, 10, 11, 12.

No se encontraron índices de discriminación negativos, que sería los que peor poder predictivo implicarían.

Aunque se trata de una medida sencilla y ampliamente utilizada del poder discriminativo del ítem, el valor d está sesgado hacia los grados intermedios de dificultad. Es decir, si un ítem es muy fácil o muy difícil, el valor d puede ser bajo aunque el ítem sea discriminante.

Por esta razón, se calculó además la correlación biserial puntual (Rbp), como otra medida que permite determinar el grado en que el constructo que se mide con el test también es medido por ítem. Proporciona una estimación de la correlación producto–momento de Pearson entre la calificación total del test y el continuo hipotético del ítem dicotómico (Henrysson, 1971). Por lo que es otra medida de la relación entre el ítem y el criterio, independiente de la dificultad que presente el ítem. Para Guilford y Frutchter (1978), los valores de esta correlación pueden oscilar entre -1 y $+1$, pasando por cero (ausencia de correlación).

En la presente muestra, los valores de la correlación biserial–puntual se encontraron entre .05 y .517. Es decir, todos los ítem del test presentaron correlaciones positivas y significativas con el criterio, a excepción del ítem E11 (Ver Tabla 3).

Tabla 3.

Proporción de respuesta por distractor, Índice de Dificultad, Discriminación y Correlación biserial-puntual. Nota. * $p < .05$; ** $p < .01$; *** $p < .001$ (bilateral)

Serie	Item	Opciones de respuesta ¹							
		1	2	3	4	5	Índice Dificultad (p)	Índice Discriminación ² (d)	Correlación Biserial-puntual ³ (Rbp)
B	1	.021	0	0	0	.949	.963	0.097	.253**
	2	.03	0	.003	.938	0	.954	0.101	.196**
	3	.94	0	.015	.006	.009	.964	0.101	.256**
	4	.003	.015	.006	.027	.914	.892	0.274	.349**
	5	.009	.935	.003	.006	.015	.932	0.190	.343**
	6	.866	0	.009	.003	.089	.859	0.298	.357**
	7	0	.003	.024	.086	.857	.826	0.387	.422**
	8	.006	.024	.911	.018	.012	.898	0.270	.377**
	9	.018	.018	.012	.899	.024	.909	0.169	.290**
	10	.003	0	.938	.015	.015	.940	0.177	.324**
	11	.003	.003	.077	.836	.051	.825	0.419	.439**
	12	.848	.009	.024	.057	.033	.822	0.496	.502**
C	1	.042	.021	.042	.036	.827	.779	0.508	.484**
	2	.003	.110	.033	.815	.009	.750	0.512	.470**
	3	.042	.128	.021	.030	.747	.721	0.456	.403**
	4	.685	.051	.036	.182	.018	.635	0.613	.484**
	5	.030	.149	.744	.036	.012	.716	0.423	.381**
	6	.054	.006	.643	.241	.027	.616	0.581	.455**
	7	.241	.095	.482	.018	.131	.457	0.605	.454**
	8	.688	.071	.119	.080	.012	.634	0.524	.435**
	9	.292	.307	.089	.170	.110	.284	0.375	.363**
	10	.095	.405	.173	.092	.202	.350	0.565	.446**
	11	.256	.086	.116	.146	.360	.177	0.008	.083*
	12	.057	.113	.214	.202	.378	.103	0.206	.331**

D	1	.048	.042	.045	.789	.048	.772	0.391	.369**
	2	.054	.423	.378	.042	.074	.403	0.476	.378**
	3	.033	.417	.065	.411	.045	.402	0.585	.462**
	4	.188	.051	.232	.369	.131	.350	0.399	.347**
	5	.065	.051	.738	.083	.033	.700	0.560	.471**
	6	.048	.649	.036	.104	.134	.585	0.641	.511**
	7	.116	.045	.074	.125	.607	.526	0.492	.395**
	8	.063	.470	.140	.039	.259	.443	0.516	.405**
	9	.048	.226	.515	.128	.051	.449	0.669	.517**
	10	.083	.128	.122	.232	.402	.363	0.581	.483**
	11	.113	.211	.071	.104	.467	.384	0.569	.452**
	12	.036	.289	.256	.256	.131	.239	0.339	.357**
E	1	.256	.152	.265	.065	.229	.252	0.327	.311**
	2	.223	.336	.137	.086	.185	.171	0.411	.465**
	3	.211	.256	.202	.250	.045	.160	0.339	.408**
	4	.113	.158	.119	.458	.110	.170	0.085	.101**
	5	.134	.250	.149	.226	.205	.209	0.298	.306**
	6	.348	.205	.265	.104	.036	.180	0.262	.305**
	7	.122	.414	.229	.134	.063	.414	0.367	.298**
	8	.134	.229	.083	.068	.443	.085	0.052	.113**
	9	.113	.277	.146	.396	.033	.114	0.157	.231**
	10	.110	.286	.414	.116	.039	.102	0.069	.117**
	11	.110	.185	.083	.173	.414	.100	0.016	.05
	12	.238	.045	.039	.274	.366	.058	0.024	.092**

¹ Proporción por opción de respuesta.

² Índice de discriminación entre los grupos superior e inferior dividiendo la distribución (27% sup. y 27% inf.).

³ Correlación Biserial Puntual.

DISCUSIÓN

El estudio, a pesar de sus limitaciones muestrales, presenta evidencias que conforman una sólida base para interpretar los resultados.

Se trata de una muestra no probabilística de personas que accedieron a participar de manera voluntaria. La muestra debería ser equilibrada en cuanto a cantidades similares de género, edad y nivel educativo alcanzado

para que sea lo más representativa posible de la población destinataria. Si por algún modo no se alcanzara esa equiparación la representatividad de dichos elementos en la muestra se vería afectada, quedando algunos grupos sub o sobre representados.

Desde los aspectos psicométricos se ha alcanzado una consistencia interna muy adecuada ($KR-20=0,865$). Este resultado es alentador, ya que informa sobre el grado en

que los distintos ítems del test miden lo mismo. Significa la constancia de los ítems para operar sobre un mismo constructo psicológico de un modo similar; en este caso, el Raven2, en el que subyace la idea de que la persona usa el proceso cognitivo esperado al responder a los ítems (capacidad cognitiva). La resolución de las Matrices Progresivas implica que los distintos dominios del funcionamiento cognitivo como la percepción, la atención, el razonamiento inductivo, la aptitud de clasificación, la espacial, el procesamiento simultáneo, el razonamiento fluido y la memoria de trabajo son requeridos para la resolución de los ítems, que se presentan de menor a mayor complejidad.

Lo antedicho se pudo observar en los índices calculados, tanto en el índice de dificultad (que se aprecia la progresión de la dificultad de resolución de los ítem a lo largo del test), como en el de discriminación (que permite diferenciar para cada ítem a sujetos con puntajes altos y bajos en el test).

Por último es importante mencionar que en Argentina se utilizan muchos tests extranjeros que son adaptados de forma adecuada, sin embargo no todos cuentan con baremos locales.

Por lo tanto este estudio resulta pertinente para poder establecer un nivel de comparabilidad interpretando las puntuaciones al utilizar con baremos locales, ya que se tendría en cuenta la información demográfica contextualizada.

CONCLUSIONES

Debido a que es un trabajo en progreso (sólo se cuenta con el 54% del total de sujetos planificado) la muestra presenta sesgos ya que hay más mujeres que hombres, más cantidad de jóvenes y con más cantidad de participantes con una educación mayor a los 12 años.

Con respecto al estudio del instrumento se han observado propiedades adecuadas ya que, a pesar de las limitaciones mencionadas sobre las características muestrales, los ítem presentan complejidad creciente entre series, dificultad de los ítem y valores de una adecuada discriminación y consistencia interna.

Estos análisis aquí descritos estuvieron sujetos a las características muestrales, por lo que tal vez al incrementar la muestra se cuente con una mayor cantidad de personas con capacidades superiores que permitan observar puntajes totales más elevados, sobre todo en aquellos ítems de la serie E, la última serie del test.

Contar con los datos obtenidos en este trabajo resultan de gran utilidad de cara a la finalización del estudio con la muestra completa de 1680 participantes para poder realizar los análisis estadísticos con vistas a la construcción de los baremos de la región del NEA para adultos entre 19 y 64 años.

REFERENCIAS BIBLIOGRÁFICAS

- Ato, M., López-García, J. J. y Benavente, A. (2013). Un sistema de clasificación de los diseños de investigación en psicología. *Anales de Psicología*, 29(3), 1038–1059. <https://dx.doi.org/10.6018/analesps.29.3.178511>.
- Backhoff, E., Larrazolo, N. y Roses, M. (2000). Nivel de dificultad y poder de discriminación del Examen de Habilidades y Conocimientos Básicos (EXHCOBA). *Revista Electrónica de Investigación Educativa*: 2(1). <http://redie.uabc.mx/vol2no1/contenido-backhoff.html>.
- Cortada de Kohan, N. (1999). *Teorías psicométricas y construcción de tests*. Lugar.
- Cortada de Kohan, N. (2004). Teoría de Respuesta al Ítem: Supuestos Básicos. *Evaluar*, 4, 95–110. <https://blogs.ffyh.unc.edu.ar/indicadoresyevaluacion/files/2014/03/Teoria-de-Respuesta-al-item-1.pdf>.
- Díaz Rojas, P. A. y Leyva Sánchez, E. (2013). Metodología para determinar la calidad de los instrumentos de evaluación. *Educación Médica Superior*, 27(2), 269–286. Disponible en: <http://scielo.sld.cu/pdf/ems/v27n2/ems14213.pdf>.
- Guilford, J. y Frutchter, B. (1978). *Estadísticas fundamentales en psicología y educación*, 6ª ed. McGraw-Hill.
- Henrysson, S. (1971). Gathering, Analysing, and Using Data on Test Items. En R.L. Thorndike (Ed.), *Educational Measurement*. American Council on Education.
- International Test Commission (2000). *Pautas internacionales para el uso de los test*. Versión argentina. (Trad. Asociación Argentina de Estudio e Investigación en Psicodiagnóstico ADEIP). [Archivo PDF] <https://www.psicodiagnosticos.com.ar/pdf/pautas-internacionales.pdf>.
- International Test Commission (2017). *The ITC guidelines for translating and adapting tests*. 2nd ed. [Archivo PDF]. www.InTestCom.org.
- Kohan Cortada, A., Spada, G., Martínez, C., Rosental, S. y Carballo, N. (2024). Test de Raven en Iberoamérica. Una revisión sistemática de sus baremos. *Conexiones I*(10), 48–58. Disponible en <https://ri.conicet.gov.ar/handle/11336/233428>.
- Kuder, G. F. y Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika*, 2, 151–160.
- Lord, F. (1980). *Applications of item response theory to practical testing problems*. Hillsdale: Erlbaum Associates.
- Lloret-Segura, S., Ferreres-Traver, A., Hernández-Baeza, A. y Tomás-Marco, I. (2014). El análisis factorial exploratorio de los ítems: una guía práctica, revisada y actualizada. *Anales de psicología*, 30(3), 1151–1169. <http://dx.doi.org/10.6018/analesps.30.3.199361>.
- Pearson, Inc. (2019). *Raven's 2. Matrices Progresivas. Manual*. NCS Pearson Incorporated.
- Rasch, G. (1963). *Probabilistic models for some intelligence and attainment tests*. Denmarks Paedagogiske Institut.
- Raven, J. C., Court, J. y Raven, J. (1993). *Test de Matrices Progresivas: Escalas Coloreada, General y avanzada. Manual*. Paidós.

Artículo recibido: 18/10/2024

Artículo aceptado: 16/12/2024