

Test Raven2. Análisis preliminares para la construcción del baremo de adultos entre 19 y 64 años de las provincias del Nordeste Argentino

Ana Kohan Cortada^{1,2}, Daniela Elorrieta¹, Gabriela Spada¹, Cecilia Martínez¹, Dorita Ceberio Macías¹, Rosiel Pardo¹

¹Instituto de Investigaciones Científicas (IDIC). Universidad de la Cuenca del Plata (UCP)

² Centro interdisciplinario de investigaciones en Psicología Matemáticas y Experimental (CIIPME-CONICET)

anahancortada@conicet.gov.ar

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 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 ha administrado el **Raven2** a 917 personas de un total de 1680 participantes de las localidades de Formosa, Posadas, Corrientes, Resistencia, Sáenz Peña y zonas urbanas y rurales aledañas. La muestra está integrada por un 68% de mujeres y un 32% de hombres, con edades de $M=31,30$ y $DE= 11,23$, de distintos niveles educativos, todos participantes de la población general que han dado 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.

En este trabajo se presentan los resultados de los análisis preliminares realizados con el Test Raven2, para adultos de entre 19 y 64 años residentes en las provincias del NEA.

Palabras clave: Adultos, confiabilidad, nordeste argentino, Raven2.

Abstract

The objective of this work is to present the preliminary analyzes that will serve as a basis for the development of a local norms for the Raven's2 Test, for adults between 19 and 64 years old residing in the provinces of Northeast Argentina (NEA). This project has the financial support of Universidad de la Cuenca del Plata (UCP) and is developed by an interdisciplinary team made up of teachers, academic managers, fellows and students of Psychology, 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 widely used test in Argentina, however, it has been found that there are no norms in the NEA region (Kohan Cortada *et al.*, 2024). Due to the wide diffusion that the items of the different versions of the Raven have had over time, it is considered useful to study this new version that has all its new items generated with the Item Response Theory (IRT) to evaluate cognitive abilities. Standardized, valid and reliable tools are needed, and the foreign ones to be used must have local norms, in order to be an effective test for the evaluation of intellectual capacity and its application in the fields of clinical, educational, forensic and occupational psychology.

The team administered the Raven's2 to 917 people out of a total of 1,680 participants from the cities of Formosa, Posadas, Corrientes, Resistencia, Sáenz Peña and surrounding urban and rural areas. The sample is made up of 68% women and 32% men, with ages $M=31.30$ and $SD=11.23$, of different educational levels, all participants from the general population who have given their informed consent.

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

This work presents the results of the preliminary analyzes carried out with the Raven2 Test, for adults between 19 and 64 years old residing in the provinces of Northeast Argentina.

Keywords: Adults, reliability, northeast argentina, Raven's2.

Introducción

En la evaluación psicológica es necesario utilizar herramientas estandarizadas, válidas y confiables. También para utilizar instrumentos extranjeros se deben adaptar, validar y construir baremos locales para la toma de decisiones en los distintos ámbitos: clínico, educativo, laboral o jurídico (International Test Commission, 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 trabajo es comunicar los análisis preliminares realizados que permitirán avanzar hacia la construcción del baremo. 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 respuestas 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 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&TalentAssessment (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 (Rasch, 1963; Lord, 1980).

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 este 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 (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 es 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

La población sobre la cual se obtuvo la muestra está compuesta 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 es de tipo no probabilística. Al momento la muestra está constituida por 917 sujetos evaluados de los 1680 planificada.

Procedimiento para la recolección de datos

El procedimiento seguido ha contemplado las normativas profesionales internacionales para la adaptación y la validación de tests usados en la práctica clínica e institucional, publicadas por la International Test Commission (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 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 superaran 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 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 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 fotografieran 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 antes de la aplicación 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.

Se calcularon los descriptivos (Media, DE, Mediana, Máximo y Mínimo) para el puntaje total obtenido por 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

Tabla 1. Género y Edad

		EDAD					Total
		19 a 24	25 a 34	35 a 44	45 a 54	55 y más	
Género	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 puntaje Total del Raven2 por género y nivel educativo dicotomizado.

Tabla 2. Puntaje Total del Raven2 por género y nivel educativo dicotomizado

		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

podrían 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), también se calculó el índice de discriminación (valor D) y la correlación biserial puntual (ítem-test) 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.

Para la escala total, se calcularon los valores mínimos y máximos alcanzados.

Resultados

A continuación, se presentan los resultados preliminares del estudio.

La muestra hasta el momento posee 917 participantes. En relación con la edad, 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, siendo 39,58% y 60,41% respectivamente.

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

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 $KR-20=0,865$, lo que evidencia una consistencia interna muy buena.

También para estudiar el funcionamiento de los nuevos ítems del Raven2 en nuestra región 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 respondió 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 0,30 y 0,70 fueron los óptimos, pues indicaron que no serían 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 (0,81 a 1): B1 al B12;
- Relativamente fácil (0,66 a 0,80): C: 1,2,3,5 y D: 1,5
- Dificultad adecuada (0,51 a 0,65): C: 4,6,8 y D: 6,7
- Relativamente difícil (0,31 a 0,50): C: 7,10; D: 2,3,4,8,9,10,11 y E:7
- Difícil (0,11 a 0,30): C: 9,11, D:12, E: 1 al 6 y 9
- Muy difícil (0 a 0,10): 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 establecen 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 ítems resultaron estadísticamente significativos con $p<0,05$ a excepción de tres ítems que pertenecen a la Serie E (10, 11 y 12). Si bien los índices de discriminación encontrados con esta muestra oscilan entre los valores ,008 y ,669. Siguiendo la clasificación propuesta por Díaz Rojas y Leyva Sánchez (2013), se puede decir que tiene:

- Excelente poder discriminativo ($d>0,35$): B: 7,11,12, C: 1 a 10, D: 1 a 11 y E: 2,7.
- Buen poder discriminativo (d entre 0.26 y 0.35): B: 4,6,8, D:12 y E: 1,3,5,6
- Regular poder predictivo (d entre 0.15 y 0.25): B: 5,9,10, C:12 y E: 9
- Pobre poder predictivo (d entre 0 y 0.14): B: 1 a 3; C: 11 y D: 4,8,10,11,12

No se encontraron índices de discriminación negativos, que serían 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 (R_{bp}), se lo incluyó como otra medida que permite determinar el grado en que la capacidad que se mide con el test también las mide el 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. Para Guilford y Frutchter, (1978) Los valores pueden oscilar entre -1 y +1, pasando por cero (ausencia de correlación)

Los valores de la correlación biserial-puntual de la muestra se encuentran entre 0,05 y 0,517. Es decir, todos los ítems del test presentaron correlaciones positivas y significativas con el criterio, a excepción del ítem E11.

En la Tabla 3 se pueden observar los valores obtenidos con esta muestra.

Tabla 3. Valores de la correlación biserial-puntual de la muestra

Serie	Ítem	Opciones de respuesta ¹					Índice Dificultad (p)	Índice Discriminación ² (d)	Correlación Biserial-puntual ³ (Rbp)
		1	2	3	4	5			
B	1	0.021	0	0	0	0.949	0.963	0.097	0.253**
	2	0.03	0	0.003	0.938	0	0.954	0.101	0.196**
	3	0.94	0	0.015	0.006	0.009	0.964	0.101	0.256**
	4	0.003	0.015	0.006	0.027	0.914	0.892	0.274	0.349**
	5	0.009	0.935	0.003	0.006	0.015	0.932	0.190	0.343**
	6	0.866	0	0.009	0.003	0.089	0.859	0.298	0.357**
	7	0	0.003	0.024	0.086	0.857	0.826	0.387	0.422**
	8	0.006	0.024	0.911	0.018	0.012	0.898	0.270	0.377**
	9	0.018	0.018	0.012	0.899	0.024	0.909	0.169	0.290**
	10	0.003	0	0.938	0.015	0.015	0.940	0.177	0.324**
	11	0.003	0.003	0.077	0.836	0.051	0.825	0.419	0.439**
	12	0.848	0.009	0.024	0.057	0.033	0.822	0.496	0.502**
C	1	0.042	0.021	0.042	0.036	0.827	0.779	0.508	0.484**
	2	0.003	0.110	0.033	0.815	0.009	0.750	0.512	0.470**
	3	0.042	0.128	0.021	0.030	0.747	0.721	0.456	0.403**
	4	0.685	0.051	0.036	0.182	0.018	0.635	0.613	0.484**
	5	0.030	0.149	0.744	0.036	0.012	0.716	0.423	0.381**
	6	0.054	0.006	0.643	0.241	0.027	0.616	0.581	0.455**
	7	0.241	0.095	0.482	0.018	0.131	0.457	0.605	0.454**
	8	0.688	0.071	0.119	0.080	0.012	0.634	0.524	0.435**
	9	0.292	0.307	0.089	0.170	0.110	0.284	0.375	0.363**
	10	0.095	0.405	0.173	0.092	0.202	0.350	0.565	0.446**
	11	0.256	0.086	0.116	0.146	0.360	0.177	0.008	0.083*
	12	0.057	0.113	0.214	0.202	0.378	0.103	0.206	0.331**
D	1	0.048	0.042	0.045	0.789	0.048	0.772	0.391	0.369**
	2	0.054	0.423	0.378	0.042	0.074	0.403	0.476	0.378**

	3	0.033	0.417	0.065	0.411	0.045	0.402	0.585	0.462**
	4	0.188	0.051	0.232	0.369	0.131	0.350	0.399	0.347**
	5	0.065	0.051	0.738	0.083	0.033	0.700	0.560	0.471**
	6	0.048	0.649	0.036	0.104	0.134	0.585	0.641	0.511**
	7	0.116	0.045	0.074	0.125	0.607	0.526	0.492	0.395**
	8	0.063	0.470	0.140	0.039	0.259	0.443	0.516	0.405**
	9	0.048	0.226	0.515	0.128	0.051	0.449	0.669	0.517**
	10	0.083	0.128	0.122	0.232	0.402	0.363	0.581	0.483**
	11	0.113	0.211	0.071	0.104	0.467	0.384	0.569	0.452**
	12	0.036	0.289	0.256	0.256	0.131	0.239	0.339	0.357**
	1	0.256	0.152	0.265	0.065	0.229	0.252	0.327	0.311**
	2	0.223	0.336	0.137	0.086	0.185	0.171	0.411	0.465**
	3	0.211	0.256	0.202	0.250	0.045	0.160	0.339	0.408**
	4	0.113	0.158	0.119	0.458	0.110	0.170	0.085	0.101**
	5	0.134	0.250	0.149	0.226	0.205	0.209	0.298	0.306**
	6	0.348	0.205	0.265	0.104	0.036	0.180	0.262	0.305**
E	7	0.122	0.414	0.229	0.134	0.063	0.414	0.367	0.298**
	8	0.134	0.229	0.083	0.068	0.443	0.085	0.052	0.113**
	9	0.113	0.277	0.146	0.396	0.033	0.114	0.157	0.231**
	10	0.110	0.286	0.414	0.116	0.039	0.102	0.069	0.117**
	11	0.110	0.185	0.083	0.173	0.414	0.100	0.016	0.05
	12	0.238	0.045	0.039	0.274	0.366	0.058	0.024	0.092**

Nota. * $p < .05$; ** $p < .01$; *** $p < .001$ (bilateral)

¹Proporción por opción de respuesta.

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

³Correlación 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 la 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 la

misma cosa. Significa la constancia de los ítems para operar sobre un mismo constructo psicológico de un modo similar, en nuestro 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 ítems 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 nuestro país 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 baremos locales, ya que se tendría en cuenta la información demográfica contextualizada.

Conclusiones

Debido a que es un trabajo en progreso (sólo contamos 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 de las características muestrales los ítems presentan complejidad creciente entre series, dificultad de los ítems y valores de una adecuada discriminación y consistencia interna. Los 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 resulta 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.

Bibliografía

Aron, E. y Aron, A. (2001). *Estadística para psicología*. Pearson Education.

Ato, M., López-García, J. J., & Benavente, A. (2013). Un sistema de clasificación de los diseños de investigación en psicología. *Anales de Psicología*, 29(3), 1038-1059.

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).

Cortada de Kohan, N. (1999). *Teorías psicométricas y construcción de tests*. Ed. Lugar.

Cortada de Kohan, N. (2004). Teoría de Respuesta al Ítem: Supuestos Básicos. *Evaluar*, 4, 95-110.

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. Recuperado de: <http://scielo.sld.cu/pdf/ems/v27n2/ems14213.pdf>

Guilford, J. y Fruchter, 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*1(10), 48-58.

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>

Rasch, G. (1963). *Probabilistic models for some intelligence and attainment tests*. Denmark: Paedagogiske Institut.

Raven, J. C., Court, J. & Raven, J. (1993). *Test de Matrices Progresivas: Escalas Coloreada, General y avanzada. Manual*. Paidós.

Pearson, Inc. (2019). *Raven's2. Matrices Progresivas. Manual*. NCS Pearson Incorporated

Tornimbeni, S., Pérez, E. y Olaz, F. (2008). *Introducción a la psicometría*. Ed Paidós.