

# PROPIEDADES PSICOMÉTRICAS DE LA PRUEBA: MATRICES PROGRESIVAS DE RAVEN, EN ESTUDIANTES DE ORIENTACIÓN\*

*Marhilde Sánchez de Gallardo\*\**

*Ligia Pirela de Faría\*\*\**

Universidad del Zulia

## RESUMEN

La Investigación se planteó como objetivo determinar las propiedades psicométricas en la administración grupal de la prueba de inteligencia estandarizada Raven, descrita como de mínimo sesgo cultural, a estudiantes de la mención Orientación. Fue un estudio descriptivo correlacional, de campo y con un diseño no experimental. Se encontró un coeficiente de confiabilidad de 0.79 mediante la división por mitades y de 0.90 consistencia interna. Se obtuvieron coeficientes positivos, moderados y significativos, a nivel de 0.01, con otras pruebas de inteligencia aplicadas como Dominó (0.61), Purdue (0.56) y Otis (0.56) determinándose la prueba como válida para medir inteligencia. La correlación con el rendimiento académico resultó en un coeficiente de 0.28 calificado como positivo bajo, aunque significativo, determinando que la prueba presenta validez de constructo. Se construyó un baremo local para ser utilizado con este grupo. Se concluye la pertinencia psicométrica de este instrumento de evaluación.

**Palabras clave:** pruebas psicológicas, pruebas de inteligencia, propiedades psicométricas, rendimiento académico.

## PSYCHOMETRIC PROPERTIES OF THE TEST: RAVEN'S PROGRESSIVE MATRICES, IN GUIDANCE STUDENTS

### ABSTRACT

The Investigation had as a main goal to determine the psychometric properties of the group administration of the Progressive Matrices standardized intelligence Raven Test, described as a minimum of cultural bias, to students of Education, mention Guidance. It was a descriptive correlational study, fieldwork with a non-experimental design. We found a reliability coefficient of 0.79 by mital division and an internal consistency of 0.90. Positive, moderate and significant coefficients in the 0.01 level were obtained with other tests used as Domino (0.61), Purdue (0.56) and Otis (0.56) which determined the validity of the test to measure the intelligence. The correlation with academic performance resulted in a ratio of 0.28 described as positive, low, but significant, concluding that the test presented a construct validity. A local scale to be used with this group was built. Another conclusion is the psychometric relevance of the tested assessment tool.

**Key words:** psychological tests, intelligence tests, psychometric properties, academic performance.

Recibido: 16/02/2009 ~ Aceptado: 30/03/2009

\* Este artículo corresponde al proyecto de Investigación: Estudio comparativo de los resultados de pruebas de inteligencia, adscrito al Consejo de Desarrollo Científico y Humanístico (CONDES) a través del Centro de Investigaciones Pedagógicas (CEDIP).

\*\* Doctora en Ciencias Gerenciales. Profesora Titular Departamento de Psicología, PPI, Nivel II. marsanchez@cantv.net

\*\*\* Profesora Titular Departamento de Investigación. PPI nivel I. ligiapirela@hotmail.com. ligia.pirela@gmail.com

## INTRODUCCIÓN

Desde sus inicios, las pruebas psicológicas y particularmente las del área intelectual, han estado inmersas en una serie de controversias. Algunos psicólogos las describen como instrumentos de gran apoyo al momento de tomar decisiones importantes en la vida de las personas, como ingresos a la universidad, a organizaciones laborales; capacitación y ascensos en empresas de todo tipo, mientras otros individuos, en su mayoría de otras profesiones, indagan y destacan las limitaciones y dificultades de las mismas.

Al referirnos a instrumentos de apoyo, los describimos como una de las fuentes de información, que complementadas con otras pruebas, en el área emocional-social y personalidad, aunado a una entrevista individual, deben utilizarse para las decisiones antes mencionadas y no como un único y exclusivo insumo para generar estos resultados de gran relevancia, que involucran integralmente la vida de las personas.

Cuando se relatan a las limitaciones y dificultades, por lo general se menciona, que están muy vinculadas a la cultura de quienes las diseñan, en cuanto a contenido de los ítemes, palabras utilizadas, instrucciones, pudiendo llegar a penalizar a aquellas personas que no pertenecen a esa cultura. También se señala que están relacionadas con la inteligencia vinculada con lo que el individuo ha sido capaz de aprender, fundamentalmente en las instituciones educativas donde se ha instruido, y no con las potencialidades de la persona, o lo que es capaz de desarrollar o aprender. Las pruebas de inteligencia tradicionales están basadas en, utilizando las palabras de Catell, 1987, Kane y Engle, 2002 (citado por Feldman, 2006), la inteligencia cristalizada, definida como acumulación de información, habilidades y estrategias que se aprenden por experiencia, aplicables a la resolución de problemas y no a la inteligencia fluida, las capacidades de procesamiento de información, razonamiento y memoria, como por ejemplo, resolver analogías, agrupar una serie de letras según cierto criterio, o recordar un conjunto de números.

Otra crítica que por lo general se enuncia, está referida a que las normas a utilizar para una prueba, fueron producto de investigaciones en

grupos de individuos diferentes a las personas a quienes se les administra el instrumento.

En base a las observaciones relatadas sobre la incidencia de la cultura en los resultados de las pruebas de Inteligencia, se han diseñado pruebas tendientes a minimizar la influencia de la cultura, las cuales según Cohen y Swerdlink (2006), tendían a ser de naturaleza no verbal, con instrucciones simples y claras, administradas en forma oral por el evaluador. También enuncian estos autores que, de manera característica, las tareas no verbales consistían en ensamblar, clasificar, seleccionar o manipular objetos, y en dibujar o identificar diseños geométricos. En este intento según Morris y Maisto (2005), Aiken (2003), se creó la prueba de matrices progresivas de Raven.

El propósito de esta investigación fue analizar la solidez psicométrica de este instrumento, de lo cual se desprenden dos aspectos clave que son la confiabilidad y validez; tal como lo plantean Cohen y Swerdlink (2006) “poner a prueba las pruebas”; así mismo, se requiere comparar el desempeño de un individuo con el de otros evaluados y con normas adecuadas, se generaron normas de referencia, para identificar la posición de cada uno respecto a otros perteneciente a un grupo similar en características como edad, educación : los estudiantes de la mención orientación de la Escuela de Educación, de La Universidad del Zulia.

## **METODOLOGÍA**

Se realizó una investigación de tipo descriptiva, correlacional, con un diseño no experimental, de campo, transeccional. Se llevó a cabo en Venezuela, con la participación de 287 estudiantes de la mención Orientación, de ambos sexos, cursantes de la totalidad de los semestres de la carrera, edad promedio de 21 años, diversos años de escolaridad en la Universidad, desde primer a décimo semestre, diferentes modalidades de ingreso a la Universidad del Zulia: vía Consejo Nacional de Universidades o C.N.U. y modalidades de cambios de carrera, equivalencias y carreras simultáneas. Vale destacar que no se discriminó en los participantes el género ni el nivel socioeconómico.

En cuanto a la técnica e instrumento de recolección de datos, se administró de manera grupal la prueba de inteligencia: RAVEN. Se leyeron con detenimiento las instrucciones, en cuanto a: cómo responder los ítems y la no existencia de tiempo límite para culminar la prueba. Se constató que los participantes comprendieran las instrucciones, mediante la ejecución de los ejemplos correspondientes.

Las otras pruebas de inteligencia Otis, Purdue y Dominó, las cuales si poseen un lapso máximo de ejecución, y se aplicaron en sesiones diferentes, cada una en una oportunidad, a fin de evitar la fatiga o el cansancio en los estudiantes de la mención orientación.

La prueba Raven, está compuesta por un cuadernillo donde aparecen las figuras a las cuales les falta una parte y el sujeto debe seleccionar entre las alternativas planteadas. Además, contempla una hoja aparte para anotar las respuestas y se evalúa con una plantilla de corrección. El puntaje es 1 ó 0, resultando como máxima puntuación, 60. No presenta límite de tiempo y arroja una única calificación de inteligencia.

Se efectuaron los procedimientos estadísticos requeridos, mediante un análisis inferencial, utilizándose el programa SPSS 10.0 para Windows. La validez de la escala fue evaluada utilizándose el coeficiente de correlación de Pearson para relacionar los resultados de la prueba Raven con la de otros instrumentos de medición de inteligencia y con el rendimiento académico de los estudiantes de la mención orientación. La confiabilidad se obtuvo utilizando el Coeficiente Alfa Cronbach y el Coeficiente de Partición por mitades, con el procedimiento de Spearman-Brown (longitudes iguales) y consistencia interna.

## **FUNDAMENTOS TEÓRICOS**

### **1. Confiabilidad y validez**

Hernández, Fernández y Baptista (2006) definen la medición como el proceso de vincular conceptos abstractos con indicadores empíricos. Dicha definición incluye dos consideraciones, la primera, desde el punto de vista empírico, se refiere a que el centro de atención es la respuesta observable y la segunda, desde la perspectiva teórica, se vincula con que

el interés se sitúa en el concepto subyacente no observable, representado por la respuesta. En el mencionado proceso, el instrumento de medición o de recolección de los datos, desempeña un papel fundamental y es descrito por los autores como el recurso que utiliza el investigador para registrar información o datos sobre las variables de interés. Así mismo, Rodríguez y Pineda (2003, citado por Pelekais, Finol, Newman y Parada, 2005), lo describen como el recurso del que se vale el investigador para acercarse a la realidad y extraer información sobre su objeto de estudio. Dichos autores también señalan su utilidad en el registro de datos objetivos o subjetivos, primarios o secundarios, que se encuentran en el contexto estudiado.

Pelekais, et al. (2005) por su parte, expresan la importancia de cumplir con el proceso técnico de confiabilidad y validez en toda prueba o instrumento de carácter científico.

Ningún instrumento de medición puede tener valor a menos que mida algo en forma consistente o confiable. Por ello, otro aspecto que debe determinarse es si las pruebas son lo suficientemente confiables o no para medir aquello para lo cual se crearon.

Chávez (2007:203) define la confiabilidad como el “grado de congruencia con que se realiza la medición de una variable”. Si en ausencia de cualquier cambio permanente en una persona (producto del crecimiento, aprendizaje, enfermedad o accidente) las calificaciones de las pruebas varían en gran medida de tiempo en tiempo o en diferentes situaciones, es probable que la prueba no sea confiable y no pueda utilizarse para explicar o realizar predicciones sobre el comportamiento de las personas. (Aiken, 2003).

Münch y Ángeles (2003:55) definen la confiabilidad como “la estabilidad, consistencia y exactitud de los resultados”, considerándola siempre como un intento por procurar que los resultados obtenidos por el instrumento sean similares si se vuelven a aplicar sobre las mismas muestras en igualdad de condiciones.

Santrock (2004) señala que una prueba estable y consistente no debe fluctuar de manera significativa debido a factores aleatorios como: cuánto

durmió la noche anterior la persona que respondió la prueba, quién es el examinador o cuál es la temperatura del lugar donde se aplica la prueba.

Según Papalia y Olds (2004) se han diseñado varios métodos estadísticos para estimar el grado de confiabilidad de las mediciones: fiabilidad test-retest, formas paralelas y de división por mitades. En la fiabilidad test-retest, se aplica la misma prueba al mismo grupo o persona más de una vez. El método de formas alternativas o paralelas consiste en diseñar instrumentos, con elevadas similitudes en formato, contenido y nivel de dificultad. El otro método, utilizado en esta investigación es el método de las dos mitades, mitades partidas o split-halves; el cual requiere una sola aplicación, especificado por Hernández et al. (2006) como la comparación entre las dos mitades del conjunto total de ítems o componentes, específicamente consiste en calcular una “ $r$ ” de Pearson entre las puntuaciones de las dos mitades equivalentes de la prueba, e implica un paso adicional: ajustar la confiabilidad utilizando la fórmula Spearman-Brown.

Así mismo, el coeficiente alfa consiste en la media de todos los posibles coeficientes por mitades, corregidos según la fórmula Spearman-Brown, la cual es un índice de consistencia interna de los reactivos, es decir, la tendencia a correlacionarse en sentido positivo entre sí. Por lo general, el test se divide en dos partes, correspondientes a las preguntas pares e impares, dicho método genera una medida de consistencia interna del test.

La confiabilidad es un elemento necesario, pero no suficiente, de una buena prueba; de manera que, además de ser confiables, las pruebas deben ser precisas de una manera razonable; es decir, deben ser válidas (Cohen y Swerdlik, 2006).

En lo que respecta a la validez, Hernández et al. (2006: 277) la definen como “el grado en el que un instrumento realmente mide la variable que pretende medir”. De dicho concepto pueden presentarse distintos tipos de evidencia, la vinculada con el contenido, la relacionada con el criterio y la referida al constructo.

La validez relacionada con el criterio, según Cohen y Swerdlink (2006) es un juicio de cuán adecuadamente puede ser utilizada la puntuación de una prueba para inferir la posición más probable de un individuo con respecto a cierta medida de interés, siendo el criterio esa medida de interés. También Hernández et al. (2006) señalan que se mide dicha validez comparándola con algún criterio externo; dicho criterio es un estándar con el que se juzga la validez del instrumento. Mientras los resultados del instrumento de medición se relacionen más al criterio, la validez de criterio será mayor.

Se exponen dos tipos de evidencia de la validez relacionada con el criterio: La validez concurrente y la predictiva. La validez concurrente, exponen Cohen y Swerdlink (2006) consiste en un índice del grado en que se relaciona la puntuación de una prueba con alguna medida de criterio obtenida al mismo tiempo que la puntuación. Las declaraciones de validez concurrente indican el grado en que las puntuaciones de una prueba pueden servir para estimar la posición actual de un individuo frente a un criterio. Una vez que se ha establecido la validez de inferencia de las calificaciones, la prueba puede proporcionar una forma más rápida y menos costosa para ofrecer una decisión de clasificación o un diagnóstico, lo cual la hace atractiva para futuros usuarios.

En cuanto a la validez predictiva, para los autores Cohen y Swerdlink (2006) consiste en un índice del grado en que las puntuaciones de una prueba pronostican alguna medida de criterio. Hernández et al. (2006) la definen cuando el criterio se fija en el futuro.

Según Feldman (2006) la validez y la confiabilidad de una prueba son requisitos para una evaluación precisa de inteligencia, así como para cualquier otra medición que realicen los psicólogos. Reseña que suponiendo que una prueba es válida y confiable, se necesita interpretar el significado de la puntuación: el establecimiento de normas. Las normas son criterios de desempeño en las pruebas que permiten comparar la puntuación de una persona en una prueba, con las puntuaciones de otras personas a quienes se les administró la misma prueba. Aquellas pruebas para las que se han desarrollado normas, se denominan pruebas estandarizadas.

El autor plantea que para el desarrollo de las normas, los diseñadores de pruebas calculan la puntuación promedio de un determinado grupo para el cual se elaboró el instrumento y determinan en que grado difiere la puntuación de cada persona con respecto a las puntuaciones de los demás que se sometieron anteriormente a la prueba, de manera que, las personas evaluadas, están en posibilidades de considerar el significado de sus puntuaciones naturales en relación con la de otros que también hicieron la prueba, lo cual brinda una idea cualitativa de su desempeño.

Al respecto, es importante destacar que las muestras de las personas evaluadas que se emplean en el establecimiento de normas son cruciales para el proceso de normalización, tomando en cuenta el valor de que la gente, cuyos valores sirven para determinar las normas, debe ser representativa de los individuos a quienes se dirige la prueba.

## 2. Teorías acerca de la Inteligencia

Desde que la psicología existe como disciplina, los psicólogos han tenido diferencias acerca del mejor método para medir la inteligencia.

Santrock (2004) refiere que el problema de la inteligencia es la dificultad para observarla y medirla en forma directa, ya que sólo puede evaluarse de manera indirecta, al estudiar y comparar los actos inteligentes realizados por las personas.

Existen diferentes teorías sobre inteligencia, las cuales influyen significativamente en el contenido de los tests de inteligencia. Algunos estudiosos la describen como una aptitud general y otros como compuesta de muchas aptitudes o habilidades independientes.

Morris y Maisto (2005) plantean las consideraciones de Charles Spearman, psicólogo inglés de inicios del siglo pasado, quien sostenía que la inteligencia era bastante general, reseñada como un manantial o fuente de energía mental que fluye en la totalidad de las acciones. Al respecto, señalaba que los individuos buenos en un área, frecuentemente destacan en otras áreas. Una persona inteligente comprende con rapidez, toma decisiones acertadas, sostiene conversaciones interesantes y en general, se comporta inteligentemente en una amplia variedad de situaciones.

Spearman (1936) indicaba que cada individuo puede ser más rápido en unas áreas que en otras, producto de la manera como la misma inteligencia general subyacente se revela en diferentes actividades.

Difiriendo con Spearman, el psicólogo Thurstone (1938) argumenta que la inteligencia está integrada por siete habilidades mentales distintas: habilidad espacial, rapidez perceptual, habilidad numérica, significado verbal, memoria, fluidez verbal y razonamiento. A diferencia de Spearman, creía que estas habilidades eran relativamente independientes entre sí.

Cohen y Swerdlik (2006) describen que posteriormente, el psicólogo Cattell en 1971 en contraste con Thurstone (1938), planteó la existencia de dos grupos de habilidades mentales únicamente: Inteligencia cristalizada e Inteligencia fluida. La inteligencia cristalizada (Gc) la constituyen habilidades y conocimientos adquiridos, dependientes de la exposición a una cultura particular, así como a la educación formal e informal. La inteligencia fluida (Gf), integrada por habilidades no verbales, relativamente independientes de la cultura, así como de cualquier tipo de instrucción específica. Dicha teoría ha recibido gran atención por parte de los investigadores y usuarios de pruebas.

Más recientemente, Robert Sternberg en 1986, propuso una teoría triárquica de la inteligencia humana, en la cual señala que la inteligencia se manifiesta en tres formas principales: analítica, creativa y práctica. La primera es descrita como la capacidad para analizar, juzgar, evaluar, comparar y contrastar; la creativa se refiere a las habilidades para crear, diseñar, inventar, originar e imaginar, y la práctica, abarca las habilidades para utilizar, aplicar, implementar y poner en práctica. De manera que, algunos estudiantes son capaces en las tres áreas, otros sólo en uno o dos. Los estudiantes con patrones triárquicos diferentes, se comportan de manera distinta en las instituciones educativas, por lo cual resulta conveniente que las tareas asignadas por los docentes, requieran la combinación de estas habilidades, y que dichos educadores equilibren la instrucción con relación a los tres tipos de inteligencia, más explícitamente, deben brindar oportunidades a los estudiantes para aprender a través del pensamiento analítico, creativo y práctico (Santrock, 2006).

Adicionalmente, surge una influyente alternativa a la teoría de la inteligencia de Sternberg (1986, 2002), expuesta por Howard Gardner y colaboradores en 1983, denominada las inteligencias múltiples. La misma, plantea la inteligencia conformada por muchas capacidades separadas o inteligencias múltiples, cada una de las cuales es relativamente independiente de las otras, porque diferentes partes del cerebro procesan distintas clases de información y exponen ocupaciones en las que cada tipo de inteligencia se consideran ventajas.

Gardner (1993, citado por Papalia, Olds y Feldman, 2006) describe ocho inteligencias: \* Lógico-Matemática, con la capacidad para manejar números y resolver problemas lógicos y los campos y ocupaciones ciencia, negocio y medicina; \*Lingüística, con la capacidad para utilizar y entender palabras y matices de significado y los campos u ocupaciones, escribir, corregir y traducir; \*Espacial, descrita como la capacidad para orientarse en un entorno y juzgar las relaciones entre los objetos en el espacio y las ocupaciones: arquitectura, carpintería, planificación urbana; \*Musical, con capacidad para percibir y crear patrones tonales y rítmicos y las ocupaciones composición, dirección musical; \* Cinestésico-corporal, referida a la capacidad para desplazarse con precisión y los campos en que se utilizan, danza, atletismo, cirugía; \*Interpersonal, capacidad para entender y comunicarse con los demás y las ocupaciones, enseñanza, actuación, política; \* Intrapersonal, la capacidad para entenderse a sí mismo, con las ocupaciones de asesoría psicológica, psiquiatría y liderazgo espiritual, así como, \*Habilidades naturalistas, expuestas como la capacidad para observar patrones en la naturaleza y entender los sistemas naturales y hechos por el hombre, con las ocupaciones: caza, pesca, agricultura, botánica, ecología, cocina.

Posteriormente, Gardner (1999 citado por Kail y Cavanaugh, 2006) añade una novena inteligencia, la existencialista, en la cual se analizan asuntos *supremos*, como el propósito de la vida y la naturaleza de la muerte.

Al respecto, es importante destacar lo expuesto por Gardner, en cuanto a que las inteligencias lingüística, lógico-matemática y espacial, figuran en las teorías psicométricas, no así las otras seis.

Papalia, Olds y Feldman (2006) señalan que para Gardner un grado de inteligencia elevado en un ámbito, no necesariamente va acompañado de un grado de inteligencia elevado en cualquiera de los otros. Aunado a esto, Papalia y Olds (2004) enuncian que para el autor antes señalado, cada una de las ocho inteligencias puede ser destruida por un daño cerebral y además, como implican habilidades cognoscitivas únicas, se pueden mostrar de un modo exagerado en algunas personas *normales*, como en individuos con retardo mental o autismo.

Al comparar estos diferentes planteamientos sobre la inteligencia, Spearman creía que las personas diferían en cuanto al grado en que poseen la *energía mental* denominada inteligencia general. Teóricos posteriores como Thurstone (1938) y Cattell (1971) especificaron la estructura de las capacidades mentales, mientras que dos teóricos contemporáneos de gran influencia, Sternberg (1986) y Gardner (1993), presentan teorías similares con énfasis en las habilidades prácticas manifestadas en el mundo real pero difieren en asuntos básicos: Gardner se basó fundamentalmente en una aproximación por medio de un estudio de casos, y exploró el desarrollo de una inteligencia específica de una persona en particular, mientras que Sternberg mostró su ingenio en la creación de tests mentales para medir distintos aspectos de la inteligencia. Otra comparación importante se refiere a que las teorías de Spearman, Thurstone y Gardner, tienden a describir la manera como los individuos difieren en el contenido de la inteligencia, las diferentes capacidades subyacentes en la conducta inteligente; en cambio, Sternberg en sus trabajos más recientes sobre psicología cognoscitiva, ha enfatizado los procesos de pensamiento, que pueden ser comunes a todas las personas, en cuanto a como reúnen y utilizan información los humanos para resolver problemas y comportarse con inteligencia.

También Gardner (citado por Santrock, 2004) señala que su teoría así como la de Sternberg incluyen uno o dos categorías de inteligencia social. En los planteamientos de Gardner, las categorías son inteligencia interpersonal e inteligencia intrapersonal. En la teoría de Sternberg, la categoría es Inteligencia práctica. En la actualidad se insiste en otra teoría que resalta la importancia de los aspectos interpersonal, intrapersonal y práctico, denominada Inteligencia emocional, ampliamente difundida

por Goleman (1997). Este estudio reviste de especial importancia a las teorías de inteligencia, dado que determina el contenido de los tests de inteligencia y de habilidades mentales, los cuales se utilizan con frecuencia en la evaluación de las personas.

### 3. Pruebas de Inteligencia

En la actualidad todavía se evalúan las diferentes teorías. Sin obviar esta diatriba, según Kail y Cavanaugh (2006) es innegable que las personas se distinguen por su capacidad intelectual y se han diseñado muchas pruebas para medir las diferencias. Desde 1904, el ministro de instrucción pública de Francia, solicitó a dos famosos psicólogos de la época, Alfred Binet y Theodophile Simon construir un instrumento para identificar a los niños que no aprenderían en la escuela sin una enseñanza especial, para lo cual eligieron algunas tareas simples que los niños franceses de diferentes edades deberían ser capaces de realizar.

Santrock (2003) señala que las autoridades educativas deseaban reducir la masificación en las aulas llevando a colegios especiales a los niños que no podían seguir las enseñanzas impartidas en las escuelas convencionales. Kail y Cavanaugh (2006) exponen como a partir de pruebas preliminares, Binet y Simon (1905) identificaron algunos problemas que un niño normal de tres años podría resolver, que los niños normales de cuatro años podrían resolver y así sucesivamente. La edad mental de un niño se refería a la dificultad de los problemas que podían resolver correctamente. Un niño capaz de resolver problemas que un niño normal de siete años podría resolver tendría una edad mental de 7. Binet y Simon usaron la edad mental, para distinguir a los niños “brillantes”, de los “lentos”. Así un niño “brillante” tendría la edad mental de otro mayor y un niño “lento” tendría la edad mental de uno de menos edad. Ambos psicólogos confirmaron que los niños “brillantes” identificados mediante su prueba tenían más éxito en la escuela, por lo cual se considera que fue la primera prueba objetiva de inteligencia.

Según Morris y Maisto (2005) durante la década que siguió a la incursión de la Escala Binet-Simon, se publicaron numerosas adaptaciones

de Binet, siendo la más conocida la Escala de Inteligencia de Stanford-Binet, preparada en la Universidad de Stanford por L. M. Terman en 1916, quien además introdujo el término Coeficiente Intelectual o CI, para establecer un valor numérico de la inteligencia, asignando la puntuación de 100 a una persona de inteligencia promedio. Desde 1916, la escala Stanford-Binet se ha revisado en diferentes ocasiones, considerándose más apropiada para niños, adolescentes y adultos muy jóvenes, aunque puede utilizarse con gente mayor.

En la actualidad, junto a la prueba Stanford-Binet, se utilizan los instrumentos creados por David Wechsler, las pruebas WISC y WAIS, esta última de uso frecuente en adultos. Wechsler (1939) desarrolló la prueba fundamentalmente porque deseaba un instrumento que fuera más adecuado para adultos que la Stanford-Binet y obtuvo como resultado calificaciones separadas para la parte verbal y de desempeño, así como un CI global. Las pruebas antes mencionadas son pruebas individuales, en las cuales el examinador, lleva a la persona a una habitación aislada y pasa entre 30 y 90 minutos aplicando la prueba, posteriormente el examinador puede tardar luego otra hora para calificar la prueba de acuerdo con las detalladas instrucciones del manual, lo cual es una operación costosa que exige tiempo, además el criterio y la conducta del examinador puede tener una influencia considerable en la calificación.

Tomando en cuenta estas observaciones y los propósitos de la evaluación, surgieron las pruebas grupales, descritas por Cohen y Swerdlik (2006) como herramientas útiles para la exploración cuando se debe evaluar a una gran cantidad de personas de forma simultánea o en un período breve. Las utilizadas en este estudio: Raven, Otis, Dominó (D-48) y Purdue, corresponden a este tipo. Los autores las definen como una herramienta de evaluación, dado que son instrumentos o procedimientos utilizados para identificar un rasgo en particular o conjunto de ellos en un nivel amplio. Los datos obtenidos se podrían explorar con mayor profundidad mediante métodos de evaluación más individualizados.

Señalan Cohen y Swerdlik (2006) entre sus ventajas el uso eficiente del tiempo y de los recursos, los reactivos de la prueba se encuentran por lo general en un formato fácil de calificar por una persona o una máquina, los

evaluados trabajan independientemente a su propio paso; quien aplica la prueba no necesita mucha capacitación, ya que la tarea podría requerir tan sólo leer las instrucciones, medir el tiempo y supervisar a los evaluados; conjuntamente puede tener menos efecto en la puntuación del evaluado que otro, en una situación de uno a uno; son menos costosas que las individuales, han probado tener valor para propósitos de exploración, por lo general están normalizadas con base en grandes cantidades de personas y funcionan bien con sujetos que pueden leer, seguir instrucciones, tomar un lápiz y que no requieren mucha asistencia.

Según Aiken (2003), muchas pruebas de inteligencia colectivas se han diseñado específicamente con el propósito de medir la aptitud para el trabajo académico, por lo que se les designa pruebas de aptitud académica. Algunas pruebas de inteligencia en grupo tienen un enfoque más amplio que éste, pero aún así, su contenido es similar al de las medidas de habilidad académica: tienen un gran contenido de reactivos verbales, numéricos y otros de tipo escolar. A lo largo del tiempo, se han utilizado distintas pruebas con propósitos de admisión a las universidades e instituciones profesionales.

En este sentido, Morris y Maisto (2005) señalan que las pruebas de inteligencia tienen una fuerte correlación con las calificaciones escolares y son excelentes predictoras del desempeño en el trabajo, lo cual se vincula con lo planteado por Kail y Cavanaugh (2006) al señalar que en términos generales las pruebas logran predecir el éxito escolar en forma razonable. Estos últimos autores también exponen que las pruebas de inteligencia no sólo predicen bastante bien el desempeño escolar, sino también el trabajo, sobre todo en puestos de elevada complejidad.

No obstante, los usos de tests de inteligencia han dado lugar a una gran cantidad de controversias respecto a la naturaleza y significado de la inteligencia y las consecuencias personales y sociales que se confía dependen de estas pruebas. Por ejemplo, una de las críticas más reiteradas, según expone Santrock (2004) reflejan los valores de la cultura dominante, tales como la importancia del tiempo, idioma predominante entre diferentes grupos lingüísticos, culturas con experiencias en dibujos y fotografías, diversas actitudes, valores y motivación. También reseña

que la mayoría de los reactivos de las pruebas de inteligencia consisten en habilidades e información aprendida y en consecuencia, las personas provenientes de ambientes con privaciones tienen pocas posibilidades de haber podido adquirir los conocimientos que se requieren. Asimismo, Santrock (2003) indica dificultades en la aplicación, más que todo en las administradas grupalmente, ya que el examinador no puede establecer una relación personal, ni determinar el nivel de ansiedad de los participantes, pudiendo ocurrir que algunos estudiantes no entiendan las instrucciones o se distraigan con los compañeros de aplicación.

Sin embargo, a pesar de los planteamientos descritos, Aiken (2003) expone como méritos de las pruebas de inteligencia su importante contribución en la identificación de individuos situados en todos los niveles de las habilidades mentales y en el diseño de programas para tratarlos, entrenarlos y educarlos, tales como personas con retardo mental o superdotados; así como también para diferenciar las habilidades en un solo individuo.

### 3.1. Test de Matrices progresivas

Morris y Maisto (2005) describen esta prueba como una serie de diseños, cada uno con una sección faltante, donde la tarea consiste en encontrar de entre seis a ocho patrones, el que coincide con la parte faltante. La prueba implica varias relaciones lógicas y requiere discriminación. Según Aiken (2003) esta prueba, que puede administrarse en forma individual y colectiva, demanda al examinado indicar cual de las diversas figuras o diseños pertenece a una matriz dada. Fue desarrollada en Gran Bretaña, como una medida del factor de inteligencia general de Spearman, la prueba de Raven está disponible en The Psychological Corporation en forma de matrices progresivas Estándar, Coloreada y avanzada. La forma Estándar, para edades entre 6 y 80 años, incluye cinco conjuntos en blanco y negro de 12 problemas cada uno y se culmina en un lapso de 20 a 45 minutos. La forma coloreada para niños cuyas edades oscilan entre 5 y 11 años, individuos de edad avanzada y personas mental y físicamente impedidas, toma por lo general entre 15 a 30 minutos en culminarse. La forma avanzada, presenta un rango entre 11 años a la edad adulta y dura

entre 40 y 60 minutos. También indica que las normas de dicha prueba requieren reestandarización.

Aunado a esto, el autor la refiere como cercana a resultar justa para las culturas y representa un esfuerzo encomiable en la elaboración de pruebas en la que los distintos grupos culturales, obtengan calificaciones iguales, minimizando sesgos culturales de la sociedad occidental de la clase media. Morris y Maisto (2005) la califican como una prueba libre de cultura, diseñada para medir la inteligencia de personas que no pertenecen a la cultura en la cual se diseñaron y se minimizan o eliminan el uso del lenguaje tratando de restar importancia a las habilidades y valores, como por ejemplo, la necesidad de velocidad, variable entre las culturas.

Raven (1977) en el manual de aplicación del test de matrices progresivas para la medida de la capacidad intelectual, escala general, para sujetos de 12 a 65 años, lo enuncia como un instrumento que se presenta al examinado, integrado por 60 láminas de figuras geométricas abstractas lacunarias, ordenadas en cinco series denominadas A, B, C, D, E, de 12 ítems cada una, que plantean problemas de completamiento de sistemas de relaciones (matrices), en complejidad creciente y para cuya resolución el examinado debe deducir relaciones en las primeras 24 y correlaciones en las últimas 36 y en cada caso, demostrará haber descubierto la solución, si logra reconocerla entre varias inadecuadas que están presentes al pie de la lámina respectiva.

Más explícitamente, se distinguen, a grandes rasgos, dos grandes grupos de problemas. Las 24 primeras matrices, series A y B, son de tipo fuertemente gestáltico, cuando se plantean problemas de percepción de totalidades y el individuo debe integrar o *cerrar* una figura inconclusa y por tanto ha de percibir semejanzas, diferencias, simetría y continuidad de las partes en relación con la estructura o forma del todo.

Las 36 matrices últimas (series C, D y E) son sistemas de relaciones, plantean problemas de razonamiento y exigen operaciones analíticas de deducción de relaciones y correlatos, esto es, de pensar y discernir.

Raven (1977) lo describe como un test: sencillo, de amplio margen, económico, no verbal, no manual, interesante y agradable. Sencillo, ya que

el aprendizaje de su administración y evaluación es singularmente breve y no requiere del técnico, especial preparación o experiencias previas. De amplio margen, tomando en cuenta que se aplica a niños y adultos. Económico en personal, tiempo y material, ya que puede administrarse o autoadministrarse, en forma individual y colectiva a grupos de hasta cien sujetos, en tiempos que oscilan entre 30 y 60 minutos y el cómputo de los resultados. También fácil y ágil ya que se efectúa en breves minutos. Además el material de prueba sirve para numerosas aplicaciones. No verbal, tomando en cuenta que se puede aplicar a cualquier sujeto, independientemente de su idioma, educación y capacidad verbal, tanto por la índole del material, como por la respuesta que demanda del examinado. No manual al aplicársele a cualquier persona, cualquiera que sea su estado o capacidad motora. Interesante, pues mantiene activo el interés del sujeto en todo el transcurso de la prueba. Así como agradable, ya que la tarea se desenvuelve de modo simple y grato.

En el manual antes mencionado, Raven (1977) lo caracterizan por su objetivo, en un test de capacidad intelectual, habilidad mental general, en el cual se comparan formas y se razona por analogía, con independencia de los conocimientos adquiridos y en condiciones de disponer de tiempo ilimitado. Por la tarea interna que debe realizar el examinado, al descubrir la matriz de pensamiento implicada en cada figura, una tarea perceptual, de observación, comparación y razonamiento analógico. Tomando en cuenta la tarea externa a efectuar, es un test de complemento, cuando el individuo suministra su respuesta a los problemas planteados, completando las figuras lacunarias con el trozo que corresponda. Al mismo tiempo, por la forma de obtener la respuesta, es un test de selección múltiple, por sus empleos, una prueba de clasificación clínica y por sus fundamentos, un test factorial.

En el manual de Raven (1977) también se describe la prueba como el primer test factorial adaptado y graduado para las poblaciones de Latinoamérica y presentan baremos argentinos construidos en Buenos Aires, por la Dirección de Psicología Educativa y Orientación profesional de la Provincia de Buenos Aires y en Mendoza, producto del Laboratorio de Psicología experimental de la Universidad Nacional de Cuyo. En el

estudio efectuado en Buenos Aires, se obtuvo un retest de 0.83 a 0.93 de acuerdo con la edad y correlaciones Terman-Binet .86 y saturación G de 0.82.

Aunado a ésto, en el mencionado manual, se consideró el test de Dominó como un eficaz paralelo del Raven en la medición del factor G, resultado de investigaciones en la Universidad de Uruguay.

#### 4. Resultados

La confiabilidad de la prueba Raven evidenció resultados de 0.79, categorizado como correlación positiva considerable y la consistencia interna de 0.90 es interpretada como muy fuerte. Es decir, que existe correlación entre las dos mitades y los ítems de cada una de las subpruebas.

**Tabla N° 1:** Resultados de la confiabilidad de la prueba de inteligencia Raven.

Pruebas	División por mitades	Consistencia Interna
Raven	0.79	0.90

Todas las correlaciones de la prueba Raven con las otras mediciones colectivas de inteligencia resultaron positivas, moderadas y significativas. La correlación del Raven con la prueba Dominó demostró 0.61, con la prueba Otis 0.56 y con Purdue 0.56. Por lo tanto, se plantea que la prueba es válida para medir la inteligencia.

**Tabla N° 2:** Resultados de las correlaciones entre la prueba RAVEN con otras mediciones de inteligencia.

Pruebas	Otis	Purdue	Dominó
Raven	R=0.56** Sig. 0.00-	R=0.56** Sig. 0.00	R=0.61** Sig. 0.00

La correlación entre la medición del área intelectual efectuada con la prueba Raven y el rendimiento académico resultó en un coeficiente de 0.28. Dicha correlación es positiva y baja, aunque significativa. Por ello se concluye que, los resultados de la administración de la prueba Raven presentan validez de constructo.

**Tabla N° 3:** Resultados de la correlación entre la prueba RAVEN y rendimiento académico.

Variables	Rendimiento académico
Raven	R=0.28** Sig. 0.00

También se construyó un baremo a partir de los resultados alcanzados por los estudiantes de la mención orientación y dichos baremos sirven para ser utilizados posteriormente por este grupo, ubicándose en el nivel superior (>90) rango I, a los individuos que obtengan puntajes mayores a 53. En el nivel superior al término medio (70- 90), rango II a quienes revelen puntuaciones entre 48 y 53. En, término medio (30-70), rango III a quienes muestren puntajes entre 42 y 47. En el inferior al término medio (30-70) rango IV, a quienes evidencien puntuaciones entre 33 y 41 y por último, inferior (<10) rango V a puntajes menores a 33.

**Tabla N° 4:** Normas para la interpretación del nivel intelectual de los estudiantes de Educación mención Orientación.

Rango	Nivel	Raven
I	Superior (>90)	>53
II	Superior al termino medio (70- 90)	48-53
III	Termino medio (30-70)	42-47
IV	Inferior al término medio (30 – 10)	33-41
V	Inferior (<10)	<33

## CONCLUSIONES

La Inteligencia es un constructo psicológico que desde hace más de un siglo ha intentado ser cuantificado por diversos investigadores, por ello se han diseñado diferentes instrumentos de medición, generados a partir de perspectivas y propósitos heterogéneos.

En las últimas décadas, ha existido un interés por analizar las propiedades psicométricas de los test utilizados con carácter científico, con el propósito de confirmar su pertinencia en la evaluación de las capacidades intelectuales, destacando el aporte de dicha información en la toma de decisiones sobre las personas. Entre dichas pruebas, se encuentra la denominada Matrices progresivas de Raven, descrita por Morris y Maisto (2005) y Aiken (2003) con un mínimo de sesgo cultural. En el presente estudio, se administró de manera colectiva dicho instrumento a estudiantes de la mención orientación, encontrándose las siguientes evidencias:

La prueba Raven presenta elevados coeficientes de confiabilidad y consistencia interna, indicando exactitud y estabilidad en los resultados. En cuanto a la validez concurrente también muestra coeficientes de correlación moderados y significativos con respecto a otras pruebas de inteligencia, motivo por el cual se concluye que mide el constructo que dice medir; así mismo revela vinculación positiva, baja, aunque significativa con rendimiento académico; por lo que pudiese utilizarse como alternativa en la medición del constructo. Ello, tomando en cuenta el aporte de las normas locales generadas en la investigación y partiendo de las respuestas de los estudiantes de la mención orientación para la interpretación de los puntajes.

Con base en los resultados obtenidos, se sugiere: i) continuar indagando las propiedades psicométricas de dicho instrumento, al relacionar los resultados generados en la prueba con las opiniones de sus docentes; ii) administrar nuevamente el instrumento a fin de estimar la confiabilidad test-retest, iii) analizar los errores cometidos en la prueba por aquellas personas que obtuvieron menores puntuaciones así como las posibles causas que incidieron en esos bajos puntajes. Además, se

recomienda investigar otras variables que pudiesen estar relacionadas con el rendimiento académico, como por ejemplo: locus de control, motivación al logro, entre otras. Estas variables junto al coeficiente intelectual obtenido en la administración grupal antes descrita, pudieran dar una mejor explicación al rendimiento académico.

## REFERENCIAS

- Aiken, L. (2003). Test psicológicos y evaluación. Editorial Pearson, México D.F., México.
- Cohen, R. y Swerdlink, M. (2006). Pruebas y Evaluación Psicológicas. Editorial Mc. Graw Hill Interamericana, Madrid, España.
- Chávez, N. (2007). Introducción a la Investigación Educativa. Editorial ARS Gráficos. Maracaibo, Venezuela.
- Feldman, R. (2006). Psicología con aplicaciones en países de habla hispana. Mc Graw Hill Interamericana, México D.F., México.
- Gregory, R. (2001). Evaluación Psicológica. Historia, principios y aplicaciones. Editorial El Manual Moderno, México D.F. México.
- Hernández, R. Fernández, C. y Baptista, P. (2006). Metodología de la Investigación. Editorial Mc Graw Hill Interamericana. México, D.F., México.
- Kail, R. y Cavanaugh, J. (2006). Desarrollo Humano. Una perspectiva del ciclo vital. Thompson Editores, México, D.F , México.
- Morris, C. y Maisto, A. (2005). Introducción a la Psicología. Editorial Pearson México D.F., México.
- Munch, L. y Ángeles, E. (2003). Métodos y técnicas de investigación. Editorial Trillas. México, D.F., México.
- Papalia, D. y Olds, S. (2004). Psicología. Editorial Mc. Graw Hill Interamericana. México, D.F., México.
- Papalia, D. Olds, S. y Feldman, R. (2006). Psicología del desarrollo. De la Infancia a la adolescencia. Editorial Mc. Graw Hill Interamericana. México, D.F., México.

- Pelekais, C. Finol, M. Neuman, N. Parada, J. (2005). El ABC de la Investigación. Una aproximación teórico-práctica. Ediciones Astro Data, Maracaibo, Venezuela.
- Raven, J. (1977). Test de Matrices progresivas para la medida de la capacidad intelectual. Escala general. Manual. Editorial Paidós. Buenos Aires, Argentina.
- Santrock, J. (2003). Psicología del desarrollo en la infancia. Editorial Mc. Graw Hill Interamericana, México. D.F., México.
- Santrock, J. (2004). Introducción a la Psicología. Editorial Mc. Graw Hill Interamericana, México. D.F., México.
- Santrock, J. (2006). Psicología de la Educación. Editorial Mc. Graw Hill Interamericana, México. D.F., México.