

# Análisis psicométrico del Test de Matrices Progresivas Avanzadas de Raven mediante el Modelo de Tres Parámetros de la Teoría de la Respuesta al Ítem\*

Luis Miguel Escurra Mayaute & Ana Esther Delgado Vásquez  
Universidad Nacional Mayor de San Marcos, Lima, Perú

Recibido: 30 de junio del 2010 / Aprobado: 5 de agosto del 2010

*El objetivo de este estudio fue realizar el análisis psicométrico del test de Matrices Progresivas Avanzadas de Raven mediante el modelo de tres parámetros de la Teoría de Respuesta al Ítem, en una muestra conformada por 2081 estudiantes universitarios de la ciudad de Lima. La media de la edad de los alumnos evaluados fue de  $21,3 \pm 4.28$  años, el 50,9% de sexo femenino y el 49,1% de sexo masculino. Los alumnos pertenecientes a universidades estatales eran el 63,5% y los de universidades particulares el 36,5%. Los hallazgos indican que los ítems del test no presentan una adecuada calibración de acuerdo con el Modelo de Tres Parámetros. El análisis comparativo indica que solo existen diferencias estadísticas significativas por sexo pero no por tipo de universidad.*

---

Teoría de Respuesta al Ítem / ajuste / Modelo de Tres Parámetros /Raven

---

## **A psychometric analysis of the Raven Advanced Progressive Matrices Test through the Three-Parameter Model of the Item Response Theory**

*In this study we undergo the psychometric analysis of the Raven Advanced Progressive Matrices Test through the Three-Parameter Model of the Item Response Theory administered to 2081 university students in Lima, 63.5% from state-funded universities and 36.5% from private-funded universities; 50.9% females, 49.1% males with a median age of  $21.3 \pm 4.28$ . Results show that the items of the test are not adequately calibrated to the Three-Parameter Model. Student scores are not close to normal curve expectations and on a comparative analysis, there are differences due to gender and not to university funding differences (public/private).*

---

Item Response Theory /adjustment / Three-Parameter Model / Raven

---

---

\* El presente estudio fue patrocinado por el Consejo Superior de Investigaciones (CSI) de la Universidad Nacional Mayor de San Marcos, bajo el código N.o 041801051.  
Correo electrónico: lescurra@correo.ulima.edu.pe; aedvdt2@yahoo.com.

## INTRODUCCIÓN

El Test de Matrices Progresivas Avanzadas (MPA) de Raven es una prueba no verbal, que evalúa la habilidad intelectual y el razonamiento basado en figuras. Miden la habilidad para hacer comparaciones, razonar por analogía y organizar percepciones espaciales dentro de un todo relacionado sistemáticamente. Este instrumento –junto con las otras versiones de los Test de Matrices Progresivas de Raven– se diseñó como una medida del factor *g* de Spearman o inteligencia general, que estaba constituida por dos componentes identificados como habilidad educativa y habilidad reproductiva (Raven, Court & Raven, 1995; Aiken, 1996; Anastasi & Urbina, 1998; Cohen & Swerdlik, 2001; Capolan & Sacuzzo, 2006).

Estas pruebas requieren principalmente la educción de relaciones entre reactivos abstractos, que consisten en un conjunto de matrices o arreglos de diseños en renglones y columnas de los cuales se eliminó una parte. La tarea consiste en elegir la parte faltante entre las alternativas proporcionadas. Los reactivos más sencillos requieren una discriminación precisa; los más difíciles incluyen analogías, permutaciones, alternación de patrones y otras relaciones lógicas.

El test está destinado a los adolescentes y adultos cuyo rendimiento es similar o superior al promedio (Sattler, 1988; Anastasi & Urbina, 1998). Los

Test de Raven han sido objeto de una gran cantidad de trabajos de validación y de adaptación transcultural que han demostrado la existencia de propiedades psicométricas diferenciadas, tanto en el marco de la Teoría Clásica de los Test (TCT) como en el de la Teoría de Respuesta al Ítem (Jensen & Munro, 1979; Rasch, 1980; Zagar, Arbit & Friedland, 1980; Burke, 1985; O'Leary, Rusch & Guastello, 1991; Torres & Cuesta, 1992; Raven, 1999; Van der Ven & Ellis, 2000; Cikrikci-Demirtasli, 2002; Sisto, Marín Rueda & Bartholomeu, 2006).

## INVESTIGACIONES ANTECEDENTES

En una de las investigaciones realizadas en el extranjero Verguts, De Boeck y Maris (1999) evaluaron el rol de la rapidez para contestar el test de MPA de Raven. Los autores observaron que la prueba presenta un conjunto de reglas que rigen a los ítems y que son indispensables para poder resolverlos, de manera que identificando dichas reglas se conceptúa que constituyen una muestra estadística de las reglas aplicadas hasta obtener la regla de la respuesta correcta. La facilidad para contestar se vio entonces como la velocidad con la cual una persona puede generar las reglas para identificar las respuestas correctas. Los autores desarrollaron una prueba que permitió aislar dicha velocidad de muestreo de la variable, y un método para verificar si

la variable fue aislada adecuadamente. Los puntajes en dicha prueba fueron comparados con el rendimiento en la prueba de MPA. Los resultados indican que la velocidad con que las personas prueban tales distribuciones constituyó una variable importante para resolver los ítems del MPA.

Van der Ven y Ellis (2000) estudiaron la unidimensionalidad de los ítems de la Escala de Matrices Progresivas Estándar (EMPAE) de Raven. La prueba se aplicó como parte de un proyecto de investigación dedicada a la identificación de niños muy dotados. La unidimensionalidad fue evaluada por medio del modelo de Rasch que se aplicó separadamente a las series A, B, C, D y E. El modelo de Rasch fue aceptado en las series A, C y D y rechazado en las series B y E, lo cual indicó que los ítems de dichas series miden por lo menos dos dimensiones diferentes. También se asumió que estas dimensiones son continuaciones de la Gestalt y el razonamiento analógico para la serie B, y el razonamiento analógico y afrontamiento para el bloque E. En el caso de la serie C se asumió que la homogeneidad de Rasch podría ser mejorada asumiendo un segundo factor, aparte del razonamiento analógico que fue identificado como la falta de resistencia a los distractores perceptivos. Cuando se dividió la serie B en los subconjuntos apropiados se obtuvieron dos subconjuntos unidimensionales, B1 y B2. Al dividirse la serie E se obtu-

vo un bloque unidimensional E1 y un bloque heterogéneo multidimensional E2. La serie C fue redefinida eliminando algunos de sus ítems. En el nivel del recientemente redefinido subconjunto se observó el factor de razonamiento analógico como el común a todos los subconjuntos. El factor de la continuación de Gestalt es común a las series A y B1. Sin embargo, la confiabilidad de estos subconjuntos fue muy baja, lo cual implicó que este factor sería demasiado débil para que pueda ser identificado en un análisis factorial exploratorio.

Los factores de afrontamiento y de falta de resistencia a los distractores perceptuales son considerados únicos, por lo que se podría esperar la emergencia de solo un factor al realizar un análisis factorial de todos los subconjuntos de las series definidas. Sin embargo, el análisis factorial realizado de los subconjuntos permitió identificar dos factores, aunque más allá la inspección de la parcela del segundo factor demostró que su aparición fue considerada como un artefacto debido a la asimetría de los puntajes de los subconjuntos.

Bors y Vigneau (2001), por su parte, evaluaron el efecto de la práctica con el Test de Raven Avanzado, para ello trabajaron con una muestra de 77 participantes (39 hombres y 28 mujeres), con edades entre 26 y 79 años, a los cuales se les administró en tres ocasiones el MPA. Los autores reportaron que si

bien los puntajes de los MPA totales fueron muy confiables en las tres ocasiones, las confiabilidades de la mayoría de los ítems fueron sumamente bajas. Además, reportaron que el modelo de un solo factor fue un modelo adecuado, pues explicaba el 20% de la varianza total para la matriz de las correlaciones inter-ítem en las tres ocasiones. También se encontró que el total de los puntajes del MPA aumentó significativamente en las tres ocasiones (aproximadamente dos ítems por cada ocasión). Los autores reportaron que las mejoras en el puntaje total por las ocasiones ocurrieron en un contexto en el cual los sujetos corrigieron las respuestas incorrectas de la ocasión anterior. El número de ítems sin contestar no se relacionó con ambos MPA en cualquier ocasión dada y la cantidad de ganancia en el puntaje realizado en las ocasiones. Estos resultados sugirieron que las mejoras en el desempeño no estuvieron basadas en la adquisición de un plan estratégico para responder a más ítems o en la retención de la información de un ítem específico, sino que la mejora reflejó un aprendizaje de algún aspecto común en los tipos de ítems encontrados en el MPA.

Babcock (2002) realizó un estudio donde examinó las diferencias de edades adultas en los tipos de errores realizados en el test de MPA de Raven (MPA). La muestra total fue de 818 casos, los cuales fueron compilados en tres estudios mayores en los que se

aplicó el MPA. Se determinaron los tipos de errores que los participantes cometieron, los distractores para cada problema del MPA fueron primero codificados como uno de cuatro errores posibles: correlato incompleto (IC), principio mal aplicado (WP), confluencia de ideas (CF) y repetición (RP).

Los puntajes de cada tipo de error fueron calculados para cada participante de manera individual y reflejó la desviación de seleccionar por casualidad exclusivamente un tipo de error. Los análisis realizados compararon tres grupos de edad (joven, medio y adulto) y tres niveles de habilidad (bajo, medio y alto). Los resultados indicaron que los adultos de todas las edades cometieron tipos similares de errores, pero que los adultos de niveles de habilidades diferentes cometieron distintos tipos de errores. También se examinaron la edad y las diferencias de habilidad en los problemas por el tipo de reglas requeridas, como lo clasificó Carpenter; los resultados permitieron apreciar que a medida que se incrementaban la edad y la habilidad se presentaron diferencias en el número de errores cometidos cuando se involucraron diferentes tipos de reglas.

Rushton, Skuy y Fridjhon (2003) analizaron si el MPA tenía la misma validez de constructo en los estudiantes universitarios africanos y en los no africanos. Los análisis fueron realizados en 294 estudiantes, de entre 17 y 23 años, pertenecientes a las facultades

de Ingeniería y Arquitectura de la Universidad de Witwatersrand. La muestra estuvo conformada por 187 africanos, 40 indios orientales y 67 alumnos blancos y la composición por sexo fue de 70 mujeres y 224 hombres. La aplicación del Test de Raven avanzado, indicó que hubo diferencias estadísticas significativas ( $p < .001$ ), notando que los estudiantes africanos resolvieron en promedio 22, los estudiantes indios orientales 25 y los estudiantes blancos 29. La clasificación de los valores correspondió a los puntajes de 57, 64 y 86, los cuales al traducirse en coeficientes intelectuales alcanzaron valores de 103, 106 y 117 de acuerdo con las normas estadounidenses de 1993 que fueron utilizadas. Los autores indicaron que cuatro meses antes habían aplicado el Test de Raven superior, al correlacionar ambas pruebas alcanzaron un valor de  $r = 0.60$ , superior para los africanos y los no africanos, aunque ambas pruebas predijeron mal el rendimiento para el último fin de año, pues solo alcanzaron correlaciones de  $r = 0.30$   $p < .05$ .

Las diferencias entre indios y blancos, y africanos y orientales fueron clasificadas como producidas por el denominado efecto Jensen, el cual fue más pronunciado en el factor general de inteligencia, que fue medido en los ítems a través de las correlaciones ítem-total más altas. Las cargas de  $g$  presentaron una generalización transcultural, pues

las correlaciones ítem-total calculadas en los estudiantes indios orientales predijeron la magnitud de las diferencias de los africanos blancos. Adicionalmente no se encontraron diferencias estadísticas significativas por sexo.

Rushton, Skuy y Bons (2004) evaluaron la validez de constructo de las MPA, trabajando en grupos de estudiantes universitarios de ingeniería africanos y no africanos. Se trabajó con 306 estudiantes de las facultades de Ingeniería y Arquitectura de la Universidad del Witwatersrand, de los cuales 177 eran africanos, 57 indios orientales y 72 de raza blanca, siendo 54 mujeres y 252 hombres, y cuyas edades oscilaban entre 17 y 23 años. En el desarrollo del análisis se aplicó una prueba de comprensión inglesa, la subescala de similitudes de la Escala de Inteligencia Adulta de Wechsler, el promedio de rendimiento académico al finalizar el año universitario y las notas cuando finalizó la educación secundaria. Los resultados indicaron que al aplicarse los 36 ítems del Test de Raven avanzado se encontraron diferencias estadísticas significativas ( $p < .001$ ); los estudiantes africanos alcanzaron en promedio 23; los estudiantes indios orientales 26 y los estudiantes blancos 29. Al clasificarlos en percentiles se observó que correspondieron a los puntajes de 60, 71 y 86, respectivamente, que al ser transformados en coeficientes intelectuales alcanzaron

valores de 103, 108 y 118, de acuerdo con las normas norteamericanas de 1993 que fueron aplicadas.

Los investigadores reportaron que el mismo patrón de diferencias entre los grupos fue observado en la prueba de comprensión, la subescala de similitudes y los rendimientos universitarios y secundarios. Además, indicaron que los ítems del Test de Raven se comportaron de forma similar tanto en los estudiantes africanos como en los no africanos, lo que indicó la validez interna de la prueba. El análisis de los ítems y el análisis factorial confirmatorio demostraron que la diferencia entre africanos y no africanos fue más pronunciada en el factor general de inteligencia. También indicaron que la validez concurrente fue demostrada con la correlación de las matrices con las otras medidas, tanto en forma individual como compuesta. Para el grupo africano la correlación fue de 0.28 ( $p < .05$ ) y para los no africanos de 0.27 ( $p < .05$ ). Aunque el intercepto de las líneas de regresión para los dos grupos fueron significativamente diferentes, sus pendientes no lo fueron. Finalmente, se pudo concluir que los resultados del Raven avanzado fueron válidos para los africanos así como para los no africanos.

Lynn e Irwing (2004) realizaron un estudio metaanalítico, el cual incluyó 57 estudios de diferencias por sexo en muestras de poblaciones en las cuales se aplicó el Test de Matrices Progresivas

Estándar de Raven (MPS) y el Test avanzado (MPA). Los resultados indicaron que no hubo diferencias en los niños con edades entre los 6 y 14 años, pero también se observó que los varones alcanzaron valores medios superiores en la edad de 15 años. Entre los adultos, la ventaja masculina fue de 0.33, equivalente a 5 puntos de CI. Los autores concluyeron que estos resultados fueron opuestos a la creencia frecuente de que no hay ninguna diferencia por sexo en las matrices progresivas e indicaron que se apoyó en la teoría del desarrollo que propone una ventaja masculina a partir de los 15 años de edad.

Brown y Day (2006) realizaron una investigación dirigida a revisar las críticas efectuadas a las investigaciones y las metodologías de las interpretaciones de los retos de los estereotipos acerca de las diferencias entre las razas hechas con el Test MPA de Raven. Para ello se trabajó con dos grupos de estudiantes de diferentes razas (afroamericanos y blancos), a los cuales se les aplicó el MPA bajo tres condiciones según el tipo de instrucción realizada:

- a) la situación de bajo reto, donde se indicaba que se debía armar un rompecabezas,
- b) la situación de reto medio, que correspondió a las presentación de las instrucciones estandarizadas, y
- c) la situación de alto reto, que correspondió a la consigna de responder una prueba de inteligencia.

Los resultados indicaron que hubo diferencias en la interpretación del estereotipo de la diferencia de razas según los retos en los puntajes de la prueba de habilidades. Los participantes afroamericanos presentaron un bajo rendimiento respecto a los blancos en las condiciones de instrucciones de reto normal y alto, pero realizaron un desempeño superior al alcanzado por los estudiantes blancos en las instrucciones de bajo reto.

En el Perú, Delgado, Ecurra, Bulnes y Quesada (2001) adaptaron el MPA, utilizando el enfoque de la Teoría Clásica de los Test; para ello trabajaron con una muestra de 501 universitarios, los resultados indicaron que el instrumento presentó confiabilidad por consistencia interna así como validez de constructo; las comparaciones señalaron que existieron diferencias estadísticas significativas según el tipo de carrera profesional que presentaron los participantes. Hasta la actualidad el Test de MPA de Raven no se ha trabajado con la Teoría de la Respuesta al Ítem (TRI) en el contexto nacional.

### LA TEORÍA DE LA RESPUESTA AL ÍTEM (TRI)

La medición psicológica ha variado de manera considerable, debido a que ha sido influenciada por los diferentes paradigmas de investigación. Hasta hace algunas décadas imperaba el paradigma funcionalista, el cual resaltó el concepto de control experimental y que

tuvo su mayor influencia en el desarrollo de los controles, la estandarización y las normas o baremos de comparación, lo cual fue característico en la Teoría Clásica de los Test (TCT) (Allen & Yen, 2002).

En la actualidad impera el paradigma cognitivo signado por la priorización del concepto de proceso, y su influencia en el campo de la medición se ha dado sobre la base de la calibración de los ítems, la función de información y el denominado banco de ítems, lo que ha devenido en la TRI o Teoría de la Estructura Latente (Hambleton, Swaminathan & Rogers, 1991). Esto ha producido que muchas de las pruebas psicológicas construidas con el modelo de la TCT sean desarrolladas con la nueva TRI, a fin de optimizar la medición psicológica y con ello tomar mejores decisiones.

Las ventajas del enfoque de la TRI con respecto al de la TCT, de acuerdo con López Pina (1995), Embretson y Reise (2000) y Muñiz, Fidalgo, Cueto, Martínez y Moreno (2005) son las siguientes:

- a) Establecida una escala para el rasgo latente, se obtiene la misma estimación (salvo fluctuaciones del muestreo) del nivel de un sujeto cuando es medido por diferentes subconjuntos de ítem, aun cuando estos difieran en sus índices de dificultad.
- b) La estimación del nivel de habilidad de un sujeto depende solo de sus respuestas a los ítems y no de los

estadísticos descriptivos de la población de sujetos.

- c) Los parámetros de los ítems son invariantes entre diversas poblaciones de sujetos. Es decir, si la respuesta a un ítem solo depende del nivel del rasgo que se desea medir, los sujetos igualmente hábiles tendrán la misma probabilidad de respuesta correcta independientemente de la población a la que pertenecen. De este modo, ya no es necesario utilizar un grupo normativo. Sin embargo, el problema de adaptar las normas propias de la TCT se modifica, en la TRI, en el problema de verificar la invariancia de los parámetros de los ítems entre poblaciones. Si la probabilidad de contestar correctamente un ítem para un nivel dado de habilidad depende de alguna otra característica que la habilidad en cuestión, dicha probabilidad podrá variar entre las poblaciones que difieran en tal característica, con lo que el ítem resultaría sesgado al tener un funcionamiento diferencial. El funcionamiento diferencial de los ítems se presenta cuando no se satisface el supuesto de unidimensionalidad.

La TRI se caracteriza por haber desarrollado en el enfoque paramétrico un conjunto de modelos según el número de parámetros que se deben calcular; así, se tienen los modelos de 1, 2, 3 y 4 parámetros (Hambleton, 1982; Page, 1993; Muñiz, 1996).

Para De Ayala (2009) la TRI utiliza el concepto de teoría en el sentido de paradigma debido a que trata de explicar todos los hechos con los cuales puede ser confrontado (Kuhn, 1977). La TRI es un sistema de modelos que define un procedimiento para establecer la correspondencia entre las variables latentes y sus manifestaciones. No es una teoría en el sentido tradicional debido a que no explica por qué la persona presenta una respuesta específica a un ítem o por qué una persona decide cómo hacerlo.

La TRI es una teoría de la estimación estadística, utiliza una caracterización latente de los individuos y de los ítems como predictores de la respuesta observada. Muchos investigadores—como Lord (1980), Embretson (1984) y Fisher y Formann (1982)— han indicado que se utilizan las características del ítem para explicar cómo un ítem es ubicado en un punto particular. Por otra parte, la TRI alude a otros métodos de escalamiento, por ejemplo el escalograma de Guttman y los modelos de desplegamiento de Coombs, aunque los procesos cognitivos utilizados por un individuo para responder a un ítem no son configurados utilizando los modelos de la TRI.

En la TRI las personas y los ítems son ubicados en un mismo continuo. Muchos modelos suponen que la variable latente es representada por un continuo unidimensional. Adicionalmente, para un ítem existe una utilidad si

puede diferenciar entre personas localizadas a lo largo del continuo. En un ítem la capacidad de diferenciar entre personas reduce su incertidumbre acerca de su localización. Esta capacidad para diferenciar entre personas con distintas ubicaciones es mayormente una constante que puede variar entre los ítems de un instrumento.

Los análisis del rasgo latente se centran en la interacción de los sujetos con los ítems y los modelos estadísticos que recogen esta relación, los cuales se expresan gráficamente por las curvas características del ítem.

Existen dos grandes aproximaciones en la Teoría del Rasgo Latente. La diferencia entre ambas se refleja en el tipo de relación que se establece entre los datos y el modelo estadístico utilizado.

La primera de las aproximaciones se agrupa alrededor de la TRI; en ella se trata de buscar un modelo que se ajuste mejor a los datos disponibles mediante la parametrización del rasgo latente y de las propiedades de los ítems. Se les podría considerar modelos de contraste, y como proponen Thissen y Orlando (2001) los modelos de esta teoría miden lo que son los ítems, no lo que deberían ser.

La segunda aproximación corresponde a los modelos de Rasch, en los cuales se parte de un modelo determinado que dispone de unas propiedades específicas a las que los datos obtenidos se adecúan o no. Estos modelos se

denominan de ajuste; del análisis de los desajustes se obtiene la información necesaria para determinar las calibraciones de ítems y las medidas convenientes en cada caso.

La TRI (Lord, 1980) propone un modelo probabilístico que permite conocer la información proporcionada por cada ítem. De acuerdo con Hambleton, Swaminathan y Rogers (1991) el comportamiento de un sujeto evaluado ante un ítem de un test puede explicarse por el rasgo latente. Las relaciones entre la respuesta al ítem y el rasgo latente se pueden describir mediante una función denominada: Curva Característica del Ítem (CCI).

### *Supuestos básicos*

Se consideran hipótesis operativas de trabajo que serán validadas al relacionar los datos con los modelos TRI utilizados, determinando si son apropiados o no para explicar los datos (Hambleton, Swaminathan & Rogers, 1991; Muñiz, 1997; De Ayala, 2009) tal como se detallan a continuación:

- *Unidimensionalidad del rasgo latente.*- En un modelo TRI solo se mide un rasgo latente por un conjunto de ítems del test. El cumplimiento de este supuesto puede no ser exacto debido a factores tanto cognitivos como de personalidad, así como relacionados con la propia administración del test. Si existe un componente dominante que influya

en las respuestas se estará frente a un rasgo latente, en ese caso se cumple el supuesto de unidimensionalidad del modelo TRI. En caso contrario se trata de más de un rasgo latente (multidimensionalidad).

- *Independencia local estadística*.- Las respuestas a cualquier par de ítems son independientes entre sí. Ello implica que será solo el nivel del rasgo latente del evaluado el que influye en sus respuestas a los ítems del test.
- *Curva Característica del Ítem (CCI)*.- Expresa la relación entre la variable latente y las respuestas observadas al ítem; se describe como una función monótona creciente (si el nivel de rasgo latente disponible por el encuestado aumenta la probabilidad de respuesta correcta –en el caso de ítems dicotómicos– al ítem se incrementa).

Cuando el modelo es el apropiado para el conjunto de datos evaluados, y se verifican las hipótesis planteadas, se obtienen las siguientes características (Hambleton, Swaminathan & Rogers (1991):

- a) Las estimaciones del rasgo latente del sujeto no depende del test.
- b) Las estimaciones de los parámetros de los ítems no dependen de la muestra de sujetos utilizada.
- c) Los rasgos latentes estimados obtenidos de diferentes conjuntos de

ítems serán los mismos (si no se consideran los errores de medición).

- d) Los parámetros de los ítems estimados obtenidos de diferentes muestras serán los mismos (si no se consideran los errores de medición).

La TRI se caracteriza también por proporcionar estimaciones del error estándar para los rasgos individuales estimados, y no una estimación del error para todos los casos. La propiedad de invarianza de los parámetros del modelo indica que los parámetros que caracterizan un ítem no dependen de la distribución del rasgo latente de los evaluados. El parámetro que caracteriza a un evaluado no depende del conjunto de ítems del test. Se obtiene cuando el ajuste de modelo y datos es exacto en la población utilizada (Hambleton, Swaminathan & Rogers, 1991; Muñiz, 1997; De Ayala, 2009).

#### EL MODELO DE TRES PARÁMETROS

Para Hambleton (1982), Embretson y Reise (2000), Boomsma, Van Duijn y Snijders (2001), De Ayala (2009), Reckase (2009) y Nering y Ostini (2010), el Modelo de Tres Parámetros es el modelo de la TRI que se utiliza para describir las respuestas binarias (o dicotómicas) a los ítems para la medición de habilidades, los parámetros corresponden al nivel de dificultad del ítem (b), la discriminación (a) y la probabilidad de acertar por adivinación (c). Su fórmula es la siguiente:

$$P_i(\theta) = ci + \frac{1 - ci}{1 + \exp(-1.7a(\theta - bi))} \theta \in R$$

En donde

- $\theta$  : Rasgo latente: Capacidad, habilidad, atributo del sujeto encuestado.
- $P_i(\theta)$  : Probabilidad de que un encuestado elegido al azar con habilidad  $\theta$  conteste correctamente el ítem  $i$ .
- $bi$  : Parámetro de dificultad del ítem  $i$ .
- $ai$  : Parámetro de discriminación del ítem  $i$ .
- 1.7 : Valor arbitrario introducido para que la función logística se ajuste a la normal acumulada con exactitud 0.01.
- exp : Base del logaritmo neperiano: 2.178.
- $ci$  : Parámetro de pseudoazar: Probabilidad de contestar correctamente un ítem de alternativas múltiples por azar (adivinar). Para todo  $i \in \{1, 2, \dots, n\}$  (número de ítems).

El parámetro  $ci$  proporciona una posible asíntota no nula para la CCI y representa la probabilidad de respuesta correcta al ítem de los encuestados de bajo nivel del rasgo latente. Se incorpora para representar el comportamiento al más bajo nivel del continuo de la variable latente, en donde la “adivina-

ción” o pseudoazar puede ser un factor significativo ante los test de respuestas múltiples (Hambleton, Swaminathan & Rogers, 1991).

Siguiendo a Ryan (1983), se puede decir que este modelo supone que la persona al responder el ítem usa dos estrategias alternativas: su propio rasgo latente y la “adivinación”. Es difícil ver cómo un modelo que se acomoda a esas dos estrategias para contestar un ítem puede usarse para medir sujetos con un único rasgo latente unidimensional. También parece inadecuado modular la “adivinación” como una característica de los ítems del test, cuando “adivinar” parece más propio de las personas. Sin embargo, “adivinar” no parece que sea una característica ni de los ítems ni de los sujetos, ya que serán pocos los ítems que puedan ser adivinados por todos los evaluados y pocos los sujetos que adivinen todos los ítems.

## OBJETIVOS

Se plantearon los siguientes objetivos:

- Evaluar la adecuación psicométrica de los ítems que conforman el Test de MPA de Raven, de acuerdo con la TRI con base en el Modelo de Tres Parámetros.
- Establecer si los ítems del Test de MPA de Raven presentan una adecuada calibración de acuerdo con el modelo de la TRI de tres parámetros.

- Identificar si el Test de MPA de Raven presenta un adecuado ajuste al modelo de la TRI de tres parámetros.
- Evaluar la función de información que presenta el instrumento.
- Analizar si existen diferencias estadísticas significativas en los puntajes del test por tipo de universidad y por sexo.

**MÉTODO**

La presente investigación corresponde a un estudio descriptivo y psicométrico (Sánchez & Reyes, 2006; Alarcón, 2008; Hernández, Fernández & Baptista, 2010). Se trató de analizar las propiedades psicométricas del Test de MPA de Raven de acuerdo con la TRI en estudiantes universitarios de los primeros ciclos.

- *Universo y muestra de la investigación.*- El universo estuvo conformado por universitarios desde el primero hasta el cuarto año de estudios, matriculados en universidades nacionales y particulares de la ciudad de Lima. Se trabajó con una muestra representativa de 2081 casos. Para la obtención de la muestra se aplicó un procedimiento de muestreo probabilístico estratificado de acuerdo con el tamaño de las universidades seleccionadas y de las facultades.

El 63,5% de estudiantes de la muestra eran de universidades con gestión estatal y el 36,5% de universidades particulares (véase la tabla 1). Los porcentajes de los alumnos por año de estudios eran los siguientes: primer año: 27,9%;

**Tabla 1**  
**Composición de la muestra**

	Frecuencia	Porcentaje
Gestión de la universidad		
Particular	760	36,5
Estatal	1321	63,5
Año de estudios		
Primer año	580	27,9
Segundo año	530	25,5
Tercer año	501	24,1
Cuarto año	470	22,5
Sexo		
Hombre	1021	49,1
Mujer	1060	50,9

n = 2081

segundo año: 25,5%; tercer año: 24,1%, y cuarto año: 22,6%. Asimismo, el 50,1% eran mujeres y el 49,9% varones. Las edades fluctuaron entre 18 y 27 años, el promedio fue de 21 años y se obtuvo una desviación estándar de 1.24.

- *Variable de estudio.*- La variable estudiada corresponde a la medición de la inteligencia a través del Test de MPA de Raven. Adicionalmente, los resultados fueron comparados de acuerdo con el tipo de universidad de procedencia y el sexo de los alumnos.
- *Instrumento de investigación.*- Se empleó el Test de MPA de Raven, que mide la capacidad de educación de relaciones, uno de los componentes principales de la inteligencia general y del factor “g” (Raven et al., 1994; Anastasi & Urbina, 1998; Cohen & Swerdlik, 2001; Kaplan & Sacuzzo, 2006).

El instrumento se caracteriza por estar conformado por 36 matrices que deben ser completadas y organizadas con niveles de dificultad creciente.

- *Procedimiento de recolección de datos.*- Se trabajó siguiendo los siguientes pasos: En primer lugar, por cada universidad se determinaron las facultades que iban a ser evaluadas. En segundo lugar, se establecieron la cantidad de alumnos por año de estudio y las aulas

incluidas. Las evaluaciones se efectuaron de manera grupal.

- *Técnicas de procesamiento y análisis de datos.*- El análisis estadístico fue realizado tomando en cuenta las recomendaciones propuestas por Muñiz (1996), Du Toit (2003) y Linacre (2007). Inicialmente se elaboraron los manuales de codificación y calificación, la información fue procesada utilizando los paquetes estadísticos: Bilog-MG3 (Du Toit, 2003), y el Spss FWv. 18.0.

El análisis cuantitativo de los datos fue procesado en cuatro etapas, utilizándose una combinación de estadística descriptiva y procedimientos multivariados.

- Etapa I. Se realizó el análisis psicométrico del Test de MPA de Raven de acuerdo con la TCT.
- Etapa II. Se efectuó el análisis psicométrico del Test de MPA de Raven de acuerdo con la TRI.
- Etapa III. Se llevó a cabo un análisis descriptivo de las variables seleccionadas para el estudio, efectuándose el análisis de la normalidad de los puntajes a través de la prueba de Kolmogorov-Smirnov.
- Etapa IV. Se hizo el análisis inferencial, para ello se tomó en cuenta las recomendaciones propuestas por Siegel y Castellan (1995) y Catena, Ramos y

Trujillo (2003), el cual abarcó la prueba U de Mann-Whitney.

El análisis de la información se realizó con los paquetes Bilog-MG3 (Du Toit, 2003), NOHARM v. 3.0 (Fraser & Mc Donald, 2003) para el Modelo de Tres Parámetros. El Spss FW v.18 fue utilizado para los análisis psicométricos de confiabilidad, las estadísticas descriptivas, la prueba de bondad de ajuste a la curva normal de Kolmogorov-Smirnov y la prueba U de Mann-Whitney (Aron & Aron, 2001; Pagano, 2006).

Los instrumentos se aplicaron de manera grupal. Antes de la evaluación

se realizó una breve explicación de la finalidad del estudio a los alumnos con el fin de motivar su participación, dejando clara su completa libertad de participar o no.

## RESULTADOS

### *Etapa I. Análisis psicométrico mediante la TCT de la prueba de MPA de Raven*

Los resultados del análisis psicométrico con la TCT, efectuado a través del coeficiente de la correlación ítem-test corregido, presentados en la tabla 2, permiten observar que se obtienen correla-

**Tabla 2**  
**Análisis de la confiabilidad del test de MPA de Raven**

Ítem	M	D. E.	ritc	Ítem	M	D. E.	ritc
1	0.96	0.21	0.20	19	0.81	0.39	0.26
2	0.96	0.20	0.28	20	0.72	0.45	0.23
3	0.96	0.19	0.23	21	0.59	0.49	0.33
4	0.89	0.32	0.28	22	0.50	0.50	0.33
5	0.86	0.35	0.25	23	0.54	0.50	0.35
6	0.91	0.28	0.27	24	0.36	0.48	0.23
7	0.90	0.30	0.25	25	0.44	0.50	0.30
8	0.81	0.39	0.30	26	0.18	0.38	0.26
9	0.91	0.28	0.36	27	0.19	0.39	0.64
10	0.83	0.38	0.37	28	0.17	0.37	0.25
11	0.92	0.27	0.37	29	0.19	0.39	0.25
12	0.88	0.33	0.29	30	0.29	0.45	0.33
13	0.66	0.48	0.23	31	0.23	0.42	0.32
14	0.85	0.36	0.36	32	0.15	0.36	0.27
15	0.77	0.42	0.33	33	0.17	0.38	0.28
16	0.82	0.39	0.36	34	0.15	0.36	0.28
17	0.80	0.40	0.27	35	0.13	0.34	0.28
18	0.58	0.49	0.21	36	0.03	0.18	0.38

Kuder-Richardson 20 = 0.79 \*  
I C Kuder-Richardson 20 = 0.78 – 0.80

\* p < .05  
.n = 2081

ciones que fluctúan entre  $r = 0.20$  (ítem 1) y  $r = 0.64$  (ítem 27), los cuales son estadísticamente significativos ( $p < .05$ ), a la vez que superan el criterio de  $r > 0.20$  propuesto por Kline (1986), lo que permite que todos los ítems sean relevantes para la conformación de la escala. El análisis de la confiabilidad por consistencia interna, realizado a través del coeficiente Kuder-Richardson 20, alcanzó un valor de 0.79, el cual presenta un intervalo de confianza al 95%, valores entre 0.78 y 0.80, por lo que se puede concluir que el instrumento permite obtener puntajes confiables.

### ***Etapas II. Análisis psicométrico mediante la TRI de la prueba de MPA de Raven***

El análisis psicométrico mediante la TRI y el Modelo de Tres Parámetros se realizó a través de las siguientes etapas:

#### *a) Análisis de la calibración de los ítems mediante el Modelo de Tres Parámetros de la TRI*

El análisis de la calibración de los ítems del Test de MPA de Raven permite apreciar, en la tabla 3, que los ítems presentan niveles de dificultad (a) que oscilan entre un nivel muy fácil ( $b = -4.597$ ) en el ítem 1 y un nivel muy difícil ( $b = 2.286$ ) en el ítem 28. En lo que concierne a la discriminación, los valores alcanzados fluctúan entre una muy baja discriminación ( $a = 0.312$ ) en el ítem 13 y una muy alta discriminación ( $a = 2.506$ ) como en el ítem 34. En tanto la probabilidad de pseudoadvertencia alcanza valores muy bajos, que fluctúan entre  $c = 0.005$  (ítem 36) y  $c = 0.073$  (ítem 30). Además, se encuentra que los ítems que no se adecúan al Modelo de Tres Parámetros son los siguientes: 6, 10, 17, 21, 22, 23,

**Tabla 3**  
**Análisis de la calibración de los ítems del test de MPA de Raven mediante el modelo de tres parámetros**

Ítem	Parámetros			$\chi^2$	g. l.	p
	.a	.b	.c			
1	0.462	-4.597	0.029	1.60	8	.9905
2	0.784	-3.023	0.027	2.50	6	.8648
3	0.662	-3.586	0.026	6.70	7	.4619
4	0.510	-2.478	0.027	11.20	9	.2648
5	0.550	-2.581	0.022	9.70	8	.2833
6	0.579	-2.805	0.023	15.90	8	.0439
7	0.499	-2.840	0.025	12.00	9	.2147
8	0.459	-1.860	0.031	4.30	9	.8890
9	0.911	-2.031	0.034	9.50	7	.2215

(continúa)

(continuación)

Ítem	Parámetros			X <sup>2</sup>	g. l.	p
	.a	.b	.c			
10	0.880	-1.668	0.027	17.70	8	.0237
11	0.885	-2.222	0.021	7.30	7	.3991
12	0.573	-2.488	0.023	11.60	8	.1708
13	0.312	-1.785	0.022	11.60	9	.2360
14	0.811	-1.647	0.019	14.30	8	.0754
15	0.603	-1.598	0.018	12.70	9	.1786
16	0.725	-1.282	0.053	13.90	9	.1272
17	0.492	-1.920	0.018	28.20	9	.0009
18	0.355	-0.406	0.019	12.40	9	.1892
19	0.498	-2.042	0.025	10.60	9	.3046
20	0.349	-1.541	0.027	11.70	9	.2280
21	0.548	-0.670	0.016	42.30	9	.0000
22	0.554	0.140	0.021	25.50	9	.0024
23	0.577	0.130	0.021	31.60	9	.0002
24	0.456	1.289	0.015	16.60	9	.0562
25	0.530	0.310	0.010	38.40	8	.0000
26	0.473	1.217	0.036	22.90	9	.0064
27	0.647	1.860	0.021	25.90	8	.0011
28	0.601	2.286	0.030	30.30	8	.0002
29	0.819	1.919	0.045	13.30	8	.1027
30	1.256	1.178	0.073	15.80	9	.0702
31	1.562	1.299	0.063	11.10	9	.2702
32	1.402	1.553	0.044	13.70	9	.1336
33	2.450	1.378	0.069	14.40	9	.1076
34	2.506	1.432	0.046	4.90	9	.8466
35	1.907	1.534	0.030	8.40	8	.3986
36	2.188	2.237	0.005	1.90	4	.7626
M	0.844	-0.703	0.029			
D.E.	0.583	1.919	0.015			

n = 2081

**Tabla 4**  
**Análisis del ajuste del test de MPA de Raven mediante el modelo de tres parámetros**

Estadísticos	Raven avanzado
Estadístico G	542.40
Grado de libertad	299
Significación	.00001

n = 2081

24, 25, 26, 27, 28, los cuales alcanzan coeficientes chi-cuadrado que son significativos al  $p < .05$ .

*b) Análisis del ajuste al modelo de la TRI de tres parámetros*

Con la finalidad de evaluar el ajuste general del Test de MPA de Raven se calculó mediante el estadístico G, el cual alcanzó un valor de 542.40 que con 299 grados de libertad presenta una probabilidad asociada de .00001, por lo cual se concluye que existen diferencias estadísticas significativas, corroborando que el instrumento no se ajusta adecuadamente al Modelo de Tres Parámetros. (tabla 4)

*c) Análisis de la unidimensionalidad del Test de MPA de Raven mediante el Modelo de Tres Parámetros*

El análisis de la unidimensionalidad del Test de MPA de Raven fue realizado con base en el análisis factorial exploratorio no lineal, el cual se desarrolló aplicando el modelo NOHARM (Fraser &

McDonald, 2003). En la tabla 5 se incluyen las soluciones para los factores 1, 2, 3 y 4; los resultados incluidos permiten apreciar que los índices de bondad de ajuste de Tanaka varían entre 0.923 en el factor 1 y 0.989 en el factor 4, notándose además que la raíz cuadrada media de los residuales (RMSR) es menor que el valor de 0.0050 en los factores 3 y 4, de manera que los hallazgos indican que existen más de tres factores, lo cual permite concluir que el instrumento no es unidimensional.

*d) Análisis de la dependencia local condicional de los ítems del Test de MPA de Raven mediante el Modelo de Tres Parámetros*

Para analizar la dependencia local condicional de los ítems del Test de Raven se aplicó el estadístico Q3 (De Ayala, 2009). Debido a que los ítems son dicotómicos se procedió a usar la metodología de los *testlets* (Wainer, Bradlow & Wang, 2007), por lo cual se elaboraron seis agrupamientos aleatorios de seis ítems cada uno, procediéndose

**Tabla 5**  
**Análisis de la unidimensionalidad del test de MPA de Raven con el modelo HOHARM**

<b>Estadísticos</b>	<b>1 Factor</b>	<b>2 Factor</b>	<b>3 Factor</b>	<b>4 Factor</b>
Suma de cuadrados de los residuales	0.0501	0.0253	0.0106	0.0075
Raíz cuadrada media de los residuales (RMSR)	0.0089	0.0063	0.0041	0.0035
Índice de bondad de ajuste de Tanaka	0.923	0.961	0.984	0.989

N = 2081

**Tabla 6**  
**Matriz de correlaciones de los testlets del test de MPA de Raven**

	Testlet1	Testlet2	Testlet3	Testlet4	Testlet5
Testlet2	-0.0290 (0.0008)				
Testlet3	-0.0940 (0.0088)	-0.0510 (0.0026)			
Testlet4	-0.1170 (0.0137)	-0.1470 (0.0216)	-0.0130 (0.0002)		
Testlet5	-0.1570 (0.0246)	-0.1410 (0.0199)	-0.1890 (0.0437)	-0.1520 (0.0231)	
Testlet6	-0.0950 (0.0090)	-0.1360 (0.0185)	-0.1590 (0.0253)	-0.1490 (0.0222)	0.0660 (0.0044)

n = 2081

Nota:  $Q_3^2$  figura entre paréntesis.

a calcular el Modelo de Tres Parámetros, computar el coeficiente y los residuales respectivos, finalmente se correlacionaron los *testlets* entre sí. Los resultados presentados en la tabla 6 permiten observar que se encuentran valores de  $Q_3$  menores que el criterio propuesto por Yen (1993) de  $Q_3 = 0.20$ . Estos hallazgos hacen posible concluir que los *testlets* y por ende los ítems que conforman los agrupamientos no presentan dependencia local condicional.

*e) Análisis de las curvas características de los ítems del Test de MPA de Raven*

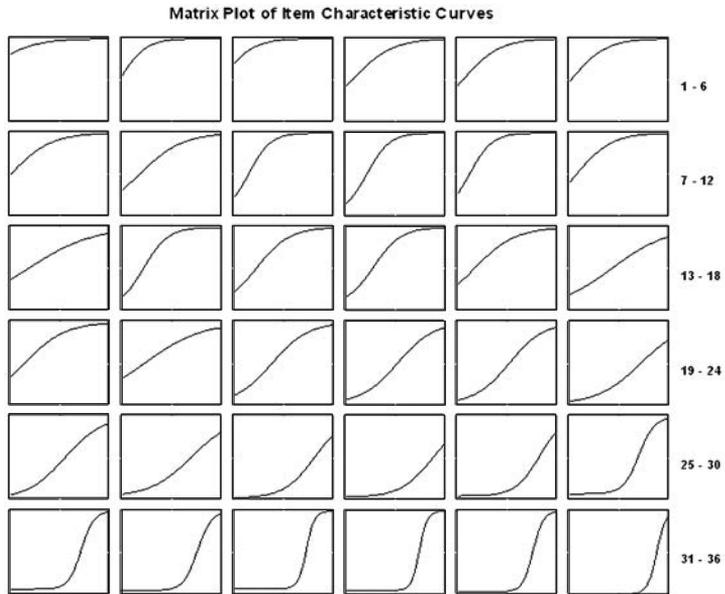
Los gráficos de las curvas características de los ítems del Test de MPA de Raven que se muestran en el gráfico 1 permiten apreciar que los ítems presentan curvas características muy diferenciadas, de tal manera que en los ítems

del 1 al 20 se observan curvas con valores más altos en el nivel de la pseudo-divinación, mientras que en los ítems del 21 al 36 presentan valores muy pequeños en el caso de la pseudo-divinación. Adicionalmente, se puede notar que las discriminaciones son más deficientes entre los ítems 21 y 35.

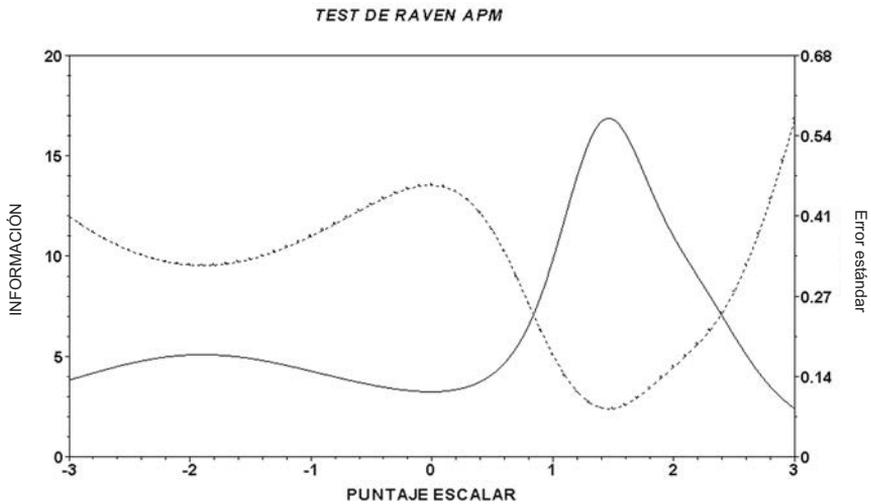
*f) Análisis de la función de información y la función de error del Test de MPA de Raven*

Los resultados que aparecen en el gráfico 2 permiten apreciar que la función de información (línea continua) no presenta una figura uniforme, sino que en los niveles bajos y medios de habilidad la información es baja, incrementándose en forma consistente en los niveles de habilidad la función de información, esto indicaría que los sujetos son mejor discriminados por el instrumento cuan-

**Gráfico 1**  
**Curvas características de los ítems del test de MPA de Raven**



**Gráfico 2**  
**Distribución de la función de información y error del test de MPA de Raven**



do tienen niveles altos de habilidad que cuando tienen niveles medios y bajos; asimismo, en lo que se refiere a la función de error, se encuentra que los niveles de habilidad bajos y medios tienen mayores niveles de error, mientras que este disminuye considerablemente cuando se evalúa a los sujetos de mayor habilidad. Estos hallazgos permiten concluir que el instrumento solo es más eficiente cuando se evalúa a sujetos con altos niveles de habilidad.

**Etapa III. Análisis descriptivo**

Los resultados del análisis de la bondad de ajuste a la curva normal fueron realizados a través de la prueba de Kol-

mogorov-Smirnov (tabla 7) e indicaron que el estadístico Z de Kolmogorov-Smirnov fue significativo (K-S Z =3.58  $p < .001$ ), por lo que se puede concluir que la distribución de los valores no se aproxima adecuadamente a la distribución normal. Es por ello que se utilizaron contrastes estadísticos no paramétricos en el análisis de los datos de la investigación, siguiendo a Siegel y Castellan (1995) y a Catena, Ramos y Trujillo (2003).

**Etapa IV. Análisis complementarios**

Para el análisis complementario se estudiaron las diferencias por sexo. Los resultados obtenidos en la tabla 8 per-

**Tabla 7**  
**Análisis de la bondad de ajuste a la curva normal de las variables estudiadas**

Variables	M	D. E.	Z de Kolmogorov-Smirnov
Raven avanzado	21.11	4.71	3.58 ***

\*\*\*  $p < 0.001$   
n = 2081

**Tabla 8**  
**Análisis del rendimiento en el test de Raven avanzado por sexo**

Variable	Varón n = 1021	Mujer n = 1060	U	Z
	Mr	Mr		
Raven avanzado	1140.30	945.36	439747.00	7.42 ***

\*\*\*  $p < .001$

**Tabla 9**  
**Análisis del rendimiento en el test de Raven avanzado por tipo de universidad**

Variable	Estatal n = 1321	Particular n = 760	U	Z
	Mr	Mr		
Raven avanzado	1123.70	1014.77	487936.00	0.634

miten apreciar que existen diferencias estadísticas significativas ( $Z = 47.42$   $p < .001$ ), notándose que los varones presentan una media de rangos mayor ( $Mr = 1140.30$ ) que las mujeres ( $Mr = 945.36$ ).

En relación con la comparación del Test de Raven avanzado por tipo de universidad (tabla 9), se puede notar que no existen diferencias estadísticas significativas ( $Z = 0.634$   $p > .05$ ) entre la universidad estatal ( $Mr = 1123.70$ ) y la universidad particular ( $Mr = 1014.77$ ).

## DISCUSIÓN

El estudio ha tenido por finalidad realizar el análisis psicométrico del Test de MPA de Raven mediante el Modelo de Tres Parámetros de la TRI. Por ello, inicialmente se hizo el análisis bajo el modelo de la TCT, y los resultados indicaron un adecuado funcionamiento psicométrico bajo este modelo, pues el índice de confiabilidad es elevado. El hallazgo concuerda con los estudios hechos por Jensen y Munro (1979),

Zagar, Arbit y Friedland (1980), Burke (1985), Kline (1986), O'Leary, Rusch y Guastello (1991), Torres y Cuesta (1992), Raven (1999) y por Delgado, Escurra, Bulnes y Quesada (2001). En conclusión, la revisión de estos resultados indica que el test funciona adecuadamente bajo los requerimientos del modelo de la TCT (Brown, 1980; Walsh & Betz, 1995; Muñoz, 1996; Muñoz, 1997; Anastasi & Urbina, 1998; Martínez, Hernández & Hernández, 2006).

Como bien indican López Pina (1995), Embretson y Reise (2000); Muñoz, Fidalgo, Cueto, Martínez y Moreno (2005); Boorsboom (2005) y De Ayala (2009), la TCT es un modelo muy laxo en sus suposiciones, de ahí que cuando se analiza bajo el modelo de la TRI mediante el Modelo de Tres Parámetros, los principales hallazgos permiten concluir que el instrumento ya no funciona en forma adecuada, hallazgos coincidentes con los estudios de Van der Ven y Ellis (2000); Cikrikci-Demirtasli (2002) y Sisto, Marín Rueda y Bartholomeu (2006).

La dificultad del Test de Raven para funcionar adecuadamente bajo el Modelo de Tres Parámetros se puede observar en los resultados de la calibración de los ítems, debido a que puede notarse que los ítems que no se adecúan estadísticamente al modelo son: 6, 10, 17, 21, 22, 23, 24, 25, 26, 27, 28. También se encuentra en el análisis global del test que tampoco se ajusta al Modelo de Tres Parámetros de la TRI.

Con relación al análisis de la unidimensionalidad del instrumento los hallazgos del Test de MPA de Raven indican que existen por lo menos dos dimensiones, tal como lo demuestran los resultados obtenidos en el análisis factorial exploratorio no lineal desarrollado con el modelo NOHARM (Fraser & Mc Donald, 2003). Estos hallazgos se corresponden con lo expuesto por Delgado, Escurra, Bulnes y Quesada (2001), quienes obtuvieron 12 factores en su estudio en ingresantes a la Universidad Nacional Mayor de San Marcos (UNMSM); asimismo, se corresponde a lo indicado por Van der Ven y Ellis (2000), quienes al estudiar la unidimensionalidad de los ítems de la escala de matrices progresivas estándar de Raven concluyó que en las series B y E medían por lo menos dos dimensiones diferentes. También concuerda con los resultados de Verguts, De Boeck y Maris (1999), los cuales reportaron que en el Test de Raven avanzado existe un factor adicional a la habilidad cognitiva y que podría influir

en los resultados, como es la rapidez para contestar el test, de manera que se presenta un conjunto de reglas que dirigen los ítems y que son indispensables para resolverlos.

Otro aspecto básico del modelo es que los ítems del instrumento presenten independencia local estocástica (estadística), es decir que no exista dependencia local condicional de los ítems considerando los resultados del estadístico Q3 (De Ayala, 2009) aplicados con la metodología de los *testlets* (Wainer, Bradlow & Wang, 2007). Debido a que los ítems son dicotómicos, y de acuerdo con el criterio de Yen (1993), se verificó que los seis *testlets* construidos y por ende los ítems que conforman los agrupamientos no presentan dependencia local condicional, lo cual corrobora que las respuestas de los ítems del test son independientes entre sí.

Respecto a la revisión de los gráficos de las curvas características de los ítems (CCI), los hallazgos indican la existencia de gráficos muy diferenciados, tanto en el nivel de las discriminaciones y las dificultades, como en los parámetros de pseudoadivinación. Adicionalmente, el análisis de la función de información y la función de error del test permite notar que en la función de información no se presenta una figura uniforme, sino que, por el contrario, predominan los niveles altos y hay un alargamiento en los niveles bajos, lo cual indicaría que el instrumento es más eficiente e informativo en los niveles de alta habilidad y menos

eficiente e informativo en los niveles de baja habilidad, este hecho es respaldado por el gráfico del error, pues el error es mayor en los niveles bajos y resulta ser mucho menor en los niveles de alta habilidad.

En lo que se refiere al análisis de la forma de la distribución de los puntajes totales, por los resultados alcanzados con la prueba de Kolmogorov-Smirnov (Siegel & Castellan, 1995; Catena, Ramos & Trujillo, 2003) se encuentra que los puntajes no se aproximan a una distribución normal, apreciándose que la distribución presenta un sesgo negativo, pues predominan los valores superiores a la media, lo que se esperaba debido al tipo de muestra utilizado en la investigación.

El análisis de las comparaciones indica que no hay diferencias por tipo de universidad, lo cual corrobora la existencia de homogeneidad en las puntuaciones en los niveles superiores. Mientras que la existencia de diferencias por sexo a favor de los varones corresponde a los hallazgos del estudio metaanalítico realizado por Lynn e Irwing (2004), quienes revisaron 57 estudios de diferencias por sexo en muestras de poblaciones en las cuales se aplicó el test de matrices progresivas estándar de Raven (SPM) y el Test avanzado (MPA). Los resultados indicaron que entre los adultos se apreció una ventaja masculina del orden de 0.33 que fue equivalente a cinco puntos de CI.; esto no coincide con la creencia de que

no hay ninguna diferencia por sexo en las pruebas de las MPA de Raven.

## CONCLUSIONES

La investigación llegó a las siguientes conclusiones:

- Los ítems del Test de MPA de Raven se adecúan mejor al modelo de la TCT.
- Los ítems no presentan una adecuada calibración de acuerdo con el modelo tres parámetros de la TRI.
- El test no presenta un adecuado ajuste al Modelo de Tres Parámetros de la TRI.
- El test no es unidimensional, el análisis con el modelo NOHARM indica que existen más de tres factores relevantes.
- El test analizado con la metodología de los *testlest*, que incluye ítems que conforman agrupamientos, no presentan dependencia local condicional.
- Los ítems presentan curvas características diferenciadas.
- El test no presenta una adecuada función de información.
- El test no presenta una adecuada función de error.
- Los varones presentan puntajes más elevados que las mujeres en el Test de MPA de Raven.
- No existen diferencias en los puntajes del test según el tipo de universidad de procedencia.

## REFERENCIAS

- Aiken, L. (1996). *Tests psicológicos y evaluación*. México: Prentice Hall-Hispanoamérica.
- Alarcón, R. (2008). *Métodos y diseños de investigación del comportamiento*. Lima: Universidad Ricardo Palma.
- Allen, M. J. & Yen, W. M. (2002). *Introduction to measurement theory*. Long Grove Illinois: Waveland Press.
- Anastasi, A. & Urbina, S. (1998). *Test psicológicos*. México: Prentice Hall.
- Aron, A. & Aron, E. (2001). *Estadística para psicología*. Buenos Aires: Pearson Education.
- Babcock, R. L. (2002). Analysis of age differences in types of errors on the Raven's Advanced Progressive Matrices. *Intelligence*, 30 (6), 485-503.
- Boomsma, A.; Van Duijn, M. A. J. & Snijders, T. A. B. (2001). *Essays on Item Response Theory*. Nueva York: Springer Verlag.
- Borsboom, D. (2005). *Measuring the mind, conceptual issues in contemporary psychometrics*. Nueva York: Cambridge University Press.
- Bors, D. A. & Vigneau, F. (2001). The effect of practice on Raven's Advanced Progressive Matrices. *Learning and Individual Differences*, 13 (4), 291-312.
- Brown, F. (1980). *Principios de la medición en psicología y educación*. México D. F.: El Manual Moderno.
- Brown, R. P. & Day, E. A. (2006). The Difference Isn't Black and White: Stereotype Threat and the Race Gap on Raven's Advanced Progressive Matrices. *Journal of Applied Psychology*, 91 (4), July, 979-985.
- Burke, H. R. (1985). Raven's Progressive Matrices (1938): More on norms, reliability and validity. *Journal of Clinical Psychology*, 41, 231-245.
- Catena, A.; Ramos, M. M. & Trujillo, H. M. (2003). *Análisis multivariado*. Madrid: Biblioteca Nueva.
- Cikrikci-Demirtasli, N. (2002). A Study of Raven Standard Progressive Matrices Test's Item measures under classic and Item Response Models: An empirical comparison. *Journal of the Faculty of Educational Sciences*, 35 (1-2), 71-79.
- Cohen, R. J. & Swerdlik, M. E. (2001). *Pruebas y evaluación psicológicas*. México D. F.: McGraw-Hill.
- De Ayala, R. (2009). *The theory and practice of Item Response Theory*. Nueva York: The Guilford Press.
- Delgado V., A.; Escurra M., L. M.; Bulnes B., M. & Quesada M., R.

- (2001). El test de matrices progresivas de Raven - forma avanzada. *Revista de Investigación Psicológica*, 4 (2), 27-40.
- Du Toit, M. (2003). *IRT from SSI*. Lincolnwood: Scientific Software International.
- Embretson, S. E. (1984). A general latent trait model for latent processes. *Psychometrika*, 49, 175-186.
- Embretson S. E. & Reise, S. P. (2000). *Item Response Theory for Psychologists*. Londres: Lawrence Erlbaum.
- Fisher, G. H. & Formann, A. K. (1982). Some applications of logistic latent trait models with linear constraints on the parameter. *Applied Psychological Measurement*, 6, 397-475.
- Fraser, C. & McDonald, R. P. (2003). *NOHARM version 3.0: A windows program for fitting both unidimensional and multidimensional normal ogive models of latent trait theor computer program*. Welland, ON: Niagara College.
- Hambleton, R. (1982). *Item Response Theory: The Tree-Parameter Logistic Model*. Los Angeles: Center for the Study of Evaluation, UCLA.
- Hambleton, R.; Swaminathan, H. & Rogers, J. (1991). *Fundamentals of Item Response Theory*. Londres: Sage Publications.
- Hernández, S. R.; Fernández, C. C. & Baptista, L. P. (2006). *Metodología de la investigación*. México D. F.: McGraw-Hill.
- Jensen, A. R. & Munro, E. N. (1979). Reaction time, movement time and intelligence. *Intelligence*, 3, 121-126.
- Kaplan, R. M. & Sacuzzo, D. P. (2006). *Pruebas psicológicas: Principios, aplicaciones y temas*. México D. F.: Thomson Learning.
- Kline, P. (1986). *Handbook of test construction*. Londres: Methuen Press.
- Kuhn, T. (1977). *The function of measurement in modern physical science*. Chicago: University of Chicago Press.
- Linacre, J. M. (2007). *A user's guide and manual to winstep, Rasch-model computer programs*. Obtenido de [www.winsteps.com](http://www.winsteps.com).
- López Pina, J. A. (1995). *Teoría de la Respuesta al Ítem: Fundamentos*. Murcia: D.M.
- Lord, F. M. (1980). *Applications of Item Response Theory to Practical Testing Problems*. Hillsdale, N.J.: Lawrence Erlbaum.
- Lynn, R. & Irwing, P. (2004). Sex differences on the progressive matrices: A meta-analysis. *Intelligence*, 32(5), 481-498.

- Martínez, M. R.; Hernández, M. J. & Hernández, M. V. (2006). *Psicometría*. Madrid: Alianza Editorial.
- Muñiz, J. (1996). *Psicometría*. Madrid: Universitas.
- Muñiz, J. (1997). *Introducción a la teoría de respuesta a los ítems*. Madrid: Universitas.
- Muñiz, J.; Fidalgo, A. M.; Cueto, E. G.; Martínez, R. & Moreno R. (2005). *Análisis de los ítems*. Madrid: La Muralla.
- Nering, M. L. & Ostini, R. (2010). *Handbook of Polytomous Item Response Theory Models*. Nueva York: Routledge.
- O'Leary, U. M.; Rusch, K. M. & Guastello, S. J. (1991). Estimating age-stratified WAIS -R IQs from scores on the Raven's Standard Progressive Matrices. *Journal of Clinical Psychology*, 47(2), 277-284.
- Pagano, R. (2006). *Estadísticas para las ciencias del comportamiento*. México, D. F.: International Thomson Editores.
- Page, M. A. (1993). *Elementos de Psicometría*. Madrid: Edudema Universidad.
- Rasch, G. (1980). *Probabilistic models for some intelligence and attainment test*. Chicago: University of Chicago Press.
- Raven, J. C. (1999). Psychometrics, cognitive ability, and occupational performance. En S. M. Wechsler & R. Souza Lobo Guzzo (Eds.). *Avaliação Psicológica: Perspectiva Internacional* (pp. 299-343). São Paulo: Casa do Psicólogo.
- Raven, J. C.; Court, J. H. & Raven, J. (1994). *Advanced Progressive Matrices, Raven Manual*. Oxford: Oxford Psychologists Press.
- Reckase, M. (2009). *Multidimensional Item Response Theory*. Nueva York: Springer Verlag.
- Rushton, J. P.; Skuy, M. & Fridjhon, P. (2003). Performance on Raven's Advanced Progressive Matrices by African, East Indian, and White engineering students in South Africa. *Intelligence*, 31, 123-137.
- Rushton, J. O.; Skuy, M. & Bons, T. A. (2004). Construct Validity of Raven's Advanced Progressive Matrices for African and Non-African Engineering Students in South Africa. *International Journal of Selection and Assessment*, 12 (3), 220-230.
- Ryan, J. (1983). Introduction to latent trait analysis and item response theory. En W. E. Hathaway (Ed.). *Testing in the schools: new directions for testing and measurement*, 19 (pp. 49-65). San Francisco: Jossey-Bass.
- Sánchez, C. H. & Reyes, M. C. (2006). *Metodología y diseños de la investigación científica*. Lima: Visión Universitaria.

- Sattler, J. (1988). *Evaluación de la inteligencia infantil y habilidades especiales*. México D. F.: El Manual Moderno.
- Siegel, S. & Castellan, N. (1995). *Estadística no paramétrica aplicada a las ciencias de la conducta*. México D. F.: Trillas.
- Sisto, F. F.; Marín Rueda, F. J. & Bartholomeu D. (2006). Estudio sobre a unidimensionalidade do Teste Matrices Progressivas Coloridas de Raven. *Psicologia: Reflexão e Critica*, 19 (1), 66-73.
- Thissen, D. & M. Orlando (2001). Item Response Theory for Items Scored in Two in Categories. En D. Thissen & H. Wainer (Eds.). *Test Scoring* (pp. 73-140). Hillsdale, NJ: Lawrence Erlbaum.
- Torres, E. & Cuesta, M. (1992). Una revisión de las propiedades psicométricas del test de Matrices Progressivas de Raven (Escala superior). *Psicothema*, 4 (1), 261-267.
- Van der Ven, A. H. G. S. & Ellis, J. L. (2000). A Rasch analysis of Raven's standard progressive matrices. *Personality and Individual Differences*, 29, 45-64.
- Verguts, T.; De Boeck, P. & Maris E. (1999). Generation speed in Raven's progressive matrices test. *Intelligence*, 27 (4), 329-345.
- Walsh, W. & Betz, N. (1995). *Tests and Assessment*. Nueva Jersey: Prentice Hall.
- Wainer, H.; Bradlow, E. T. & Wang, X. (2007). *Testlet Response Theory and its Applications*. Cambridge: Cambridge University Press.
- Yen, W. M. (1993). Scaling Performance assessments: Strategies for managing local item dependence. *Journal of Educational Measurement*, 30, 187-213.
- Zagar, R.; Arbit, J. & Friedland, J. (1980). Structure of a psychodiagnostic test battery for children. *Journal of Clinical Psychology*, 36, 313-318.

