

APLICACIÓN DE LA TEORÍA DE RESPUESTA AL ÍTEM AL TEST DE RORSCHACH

Hugo Selma Sánchez

Mario Luzardo Verde

Resumen:

El test de Rorschach es uno de los tests de referencia a nivel mundial, en especial en el contexto de la psicología clínica. Hasta hoy, el test de Rorschach ha sido analizado a través de la Teoría Clásica de los Tests (TCT). Este artículo analiza el Rorschach y las Láminas Proyectivas en una muestra representativa de 224 españoles adultos. Calculamos el grado de ajuste de algunas variables del Rorschach a uno de los modelos de la Teoría de Respuesta al Ítem (TRI), el Modelo de Respuesta Continua. Las variables que exploramos fueron: a) localización, b) movimiento humano, c) movimiento animal, d) movimiento inanimado, e) color, f) claroscuro, g) textura, h) calidad formal, i) contenido humano, j) contenido animal, k) contenidos vestimenta, alimento y sexual, l) otros contenidos y m) frecuencia. Movimiento inanimado, claroscuro y textura no pudieron ser analizadas. Localización y calidad formal no se ajustaron al modelo TRI aplicado. Movimiento humano y animal, contenidos vestimenta, alimento y sexual, y frecuencia mostraron un ajuste “adecuado” al modelo. Color, contenidos humano, animal y “otros” mostraron

un ajuste “muy bueno”. Se discuten los resultados en términos de las posibilidades y desafíos que ello implica.

Abstract:

Rorschach test is one of the psychological tests worldwide reference, especially in clinical psychology context. Since today, Rorschach test has been analyzed through the Classical Test Theory (CTT). This article analyzes Rorschach and Projective Inkblots in a sample of 224 Spanish adults. We calculated adjustment degree of some Rorschach variables to one Item Response Theory (IRT) Model, the Continuous Response Model. The variables we explore was: a) localization, b) human movement, c) animal movement, d) inanimate movement, e) color, f) chiaroscuro, g) texture, h) form quality, i) human content, j) animal content, k) clothing, food and sex content, l) other contents and m) frequency. Inanimate movement, chiaroscuro and texture can't be analyzed. Localization and form quality didn't fit to the IRT applied model. Human and animal movement, clothing, food and sex contents, and frequency, shows a “good” fit to the model. Color, human, animal and “other” contents shows a “very good” adjustment. We discuss this results in terms of possibilities and challenges that implies.

Palabras Clave: test de Rorschach, psicometría, diagnóstico psicológico, tests proyectivos, teoría de respuesta al ítem.

Keywords: Rorschach test, psychometry, psychological assessment, projective tests, item response theory.

Introducción

El test de Rorschach fue creado por el psiquiatra suizo Hermann Rorschach en el año 1921 (Rorschach, 1921). En los años siguientes fue ganando en popularidad, convirtiéndose en uno de los tests más utilizados a nivel mundial entre los años '30 y '60 (Louttit & Browne, 1947; Sundberg, 1961), muy unido a la teoría y práctica del psicoanálisis (Park, 2009). A partir de los años '70, en un contexto caracterizado por la “revolución cognitiva”, al surgimiento de la teoría humanista –que cuestionaba el uso de técnicas de evaluación- y a mayores requerimientos en cuanto a las propiedades psicométricas y a la científicidad de las técnicas psicológicas, comenzaron a surgir críticas y cuestionamientos en torno al test de Rorschach (Hunsley & Michael, 1999; Márquez Sánchez, 1986). A partir de la década de los '80, ha sido objeto de una gran controversia donde un muchos autores y asociaciones científicas y profesionales han señalado sus deficiencias psicométricas y han recomendado no utilizarlo (Hunsley & Michael, 1999; Wood, Nezworski, & Stejskal, 1996, 1997; Wood, Nezworski, Stejskal, Garven, & West, 1999). Sin embargo, gracias en parte al Sistema Comprehensivo (Exner, 1969), esa controversia parece haberse zanjado a su favor, y a día de hoy sigue siendo un test de referencia a nivel mundial (Meyer & Archer, 2001; Society for Personality Assessment, 2005), y se recomienda su utilización (Ganellen, 2001; Hibbard, 2003; Hiller, Rosenthal, Bornstein, Berry, & Brunell-Neuleib, 1999; Meyer & Archer, 2001; Meyer, Riethmiller, Brooks, Benoit, & Handler, 2000; Park, 2009; Society for Personality Assessment, 2005). Muestra propiedades psicométricas similares a otros tests ampliamente aceptados como el MMPI y el WAIS (Meyer & Archer, 2001). Se destaca además que proporciona información no obtenible mediante otros instrumentos (Society for Personality Assessment, 2005), siendo menos influenciado por variables como la deseabilidad social y la autopercepción que los ampliamente utilizados “cuestionarios de autoinforme” (Meyer & Archer, 2001). El Sistema Comprehensivo tomó los aportes de diversas escuelas que previamente habían trabajado en torno a la codificación e interpretación del test de Rorschach (Exner, 1969). Todos los códigos e índices propuestos por los principales autores precedentes, fueron sometidos a investigación empírica, con particular énfasis en los aspectos psicométricos del test (Exner, 1969). El Sistema

Comprehensive mantuvo solamente aquellos que demostraron apoyo empírico (Exner, 1969; Meyer & Archer, 2001; Park, 2009). Entre esas escuelas “previas” al sistema comprehensivo se encuentra la dirigida por Ewald Bohm (Bohm, 1973, 1979). Los datos utilizados para el presente estudio se obtuvieron mediante la aplicación y codificación mediante dicho método (Bohm, 1973). Este método, el Sistema Comprehensivo y todas las principales escuelas que han trabajado en torno al Rorschach, lo hacen a partir de un conjunto de postulados teóricos que se denomina la Teoría Clásica de los Tests o TCT (Muñiz, 2003, 2010). Dicho modelo es el más antiguo y más extendido a nivel mundial (Muñiz, 2010). El mismo posee algunas limitaciones entre las que cabe destacar (Muñiz, 2010):

- En TCT las mediciones no son independientes del instrumento a utilizar, cada instrumento tiene su *propia escala*. Por ejemplo, los puntajes en una escala de Depresión dependen de qué escala se utilicen, si se hacen 4 evaluaciones con 4 escalas distintas nosotros no podríamos afirmar qué sujeto tiene un grado mayor de depresión, porque los resultados no son comparables entre sí.
- Las propiedades de los tests en la TCT dependen directamente de los sujetos en los que se aplica el mismo. Es decir, que según el perfil de la población que estudiemos, las propiedades del test (fiabilidad, validez, etc) serán distintas. Por lo tanto... ¿las propiedades son del test, de la población, de la interacción entre ambos?
- Imprecisiones en el cálculo de la fiabilidad. La TCT asume que la fiabilidad es la misma en cualquier sujeto que realice el test, y hay abundante evidencia empírica de que esto no es así. Por ejemplo, si aplico un test de nivel de lenguaje con un gran nivel de dificultad a una persona con una afasia severa, es muy probable que falle en todos los ítems, y que algunos que acierte puedan estar influenciados por el azar... por ende. ¿Ese sujeto está siendo medido de manera fiable?

La TRI corrige estas deficiencias, y permite además una serie de

“innovaciones” que pueden resultar de gran utilidad en ámbitos clínicos y académicos. Una de ellas implica la posibilidad de comparación entre tests distintos, y entre poblaciones distintas, dejando un grupo de ítems comunes (Muñiz, 2010). Es decir, que una aplicación realizada en Japón puede compararse con una realizada en Salto, o que la aplicación de un test con un grupo de ítems a un sujeto puede ser comparada con la aplicación a otro sujeto con otro grupo de ítems, y en ambos casos bajo una *misma escala*. También está contemplada la posibilidad de creación de ítems adaptativos, es decir... que puedan adaptarse a las necesidades de evaluación: evaluar con mayor precisión a cada sujeto, evaluar con mayor precisión respecto a un determinado punto de corte, entre otras posibilidades (Muñiz, 2010).

Con el objetivo de optimizar la medición de constructos psicológicos o cognitivos y mejorar la toma de decisiones se comenzó a utilizar la Teoría de Respuesta al Ítem (TRI). La TRI toma los ítems como unidad de análisis, permite describir algunas propiedades psicométricas del instrumento mediante indicadores invariantes, es decir, que no dependen de la muestra en que se aplique. Esa se puede considerar como su mayor contribución *“la posibilidad de obtener mediciones invariantes respecto de los instrumentos utilizados y de los sujetos implicados”* (Muñiz, 1997).

La TRI propone soluciones para las limitaciones de la teoría clásica de los tests como la invarianza de los parámetros que permite que el valor de los parámetros de los ítems no depende de la muestra de donde se obtiene; la precisión con la que cada persona es medida según su nivel de rasgo y en función de los ítems concretos que se le hayan aplicado y los indicadores de bondad de ajuste que permiten estudiar el grado en que los datos ajustan al modelo (Abad, Olea, Ponsoda, & García, 2011).

La TRI se diferencia de la teoría clásica de los test por utilizar modelos basados en las características de los ítems en vez de las del test, donde las características de los ítems son independientes del grupo en que el ítem se ha calibrado y las puntuaciones del rasgo no dependen de las puntuaciones obtenidas en cada test particular. En la TRI se puede obtener una medida de la precisión para cada puntuación del rasgo, lo que la distingue claramente de la teoría clásica y para evaluar la fiabilidad no se requieren de tests estrictamente paralelos.

La teoría de respuesta al ítem establece una relación funcional entre la respuesta del examinado a cada ítem y el rasgo latente responsable de tal

realización, al que se nota θ . En la mayoría de los modelos se asume que esta función depende solo de un rasgo, es decir, son unidimensionales. La función que da la probabilidad de obtener determinada puntuación en el ítem condicionado al rasgo, se denomina curva característica del ítem (CCI).

Se supone que hay un conjunto de rasgos o aptitudes que subyacen en la respuesta de los individuos a un conjunto de ítems. Cada rasgo es un número real, por lo tanto, si existen k rasgos, estos determinan puntos del espacio R^k . A los modelos que suponen que un único rasgo es el que determina las respuestas se les llama unidimensionales. Es claro que este supuesto nunca se cumplirá estrictamente, por lo tanto, lo que se pide es que para un conjunto de ítems del test exista un factor dominante que explique la realización del test, a este factor se le llama aptitud medida por el test. Los modelos que consideran más de un factor son llamados multidimensionales.

Si se cumple el supuesto de unidimensionalidad, al aplicar a r subpoblaciones de examinados un test, las distribuciones condicionadas de las puntuaciones del test para cada nivel de rasgo son iguales para todas las subpoblaciones. En caso de no cumplirse lo anterior estamos en presencia de funcionamiento diferencial del ítem. A los efectos de la construcción de tests unidimensionales, Lumsden (1961) recomienda el uso del análisis factorial. A un conjunto inicial de ítems (construidos a partir de conocimientos previos) se les realiza un análisis factorial y se eliminan aquellos que no cargan en el primer factor. Una vez eliminados esos ítems se repite el procedimiento tantas veces como sea necesario, hasta dar con un solución satisfactoria. Este autor propone que la razón de varianzas explicada por el primer factor respecto al segundo se utilice como índice de unidimensionalidad.

Un aspecto importante es la llamada independencia local que refiere a que los n ítems que constituyen el test, condicionados a un valor constante de θ son independientes. Si tenemos un test formado por n ítems y sean U_1, U_2, \dots, U_n las variables que representan las respuestas de los examinados a los ítems si hay independencia local se cumple:

$$P(U_1 \dots U_n / \theta) = \prod_{j=1}^n P(U_j / \theta)$$

Es importante notar que la independencia local no implica que los ítems son no correlacionados, sino que son independientes condicionalmente (al nivel de rasgo). Hambleton, Swaminathan y Rogers (1991) muestran que el principio de unidimensionalidad es equivalente a la independencia local (condicionado a solo un rasgo), pues si suponemos que el test es unidimensional y mide una variable latente θ , si no se da la independencia local resultaría que algunos sujetos tendrían mayor probabilidad de responder correctamente para el mismo nivel de rasgo, lo cual niega la unidimensionalidad. Recíprocamente, si se da la independencia local tenemos que las probabilidades de respuesta solo dependen de un único rasgo. La independencia local también puede expresarse con datos multidimensionales (condicionando al espacio completo).

McDonald (1981) propone usar técnicas de análisis factorial para evaluar el supuesto de independencia local, dada la equivalencia entre independencia local y unidimensionalidad, cuando se consideran datos unifactoriales. Se define un conjunto de ítems del test como unidimensionales, cuando para sujetos con el mismo nivel de rasgo, la covariación entre los ítems es cero.

Se llama curva característica del ítem (CCI) a la función $P(\theta) : [-\infty, \infty] \mapsto [0, 1]$ que da la probabilidad de acertar el ítem condicionada a la variable latente θ , es decir,

$$P(\theta) = P(U = 1 / \theta)$$

Si el espacio latente está completamente especificado, las CCI se mantendrán invariantes para todo subgrupo de la población; luego la probabilidad que un individuo dé una respuesta correcta a determinado ítem solo depende de la forma de la curva característica del ítem y del nivel de rasgo del sujeto, y entonces, es independiente de la distribución del rasgo en la población. Esta propiedad de invarianza de las CCI (no importa en qué región fue calibrado el ítem) constituye una de las mejores características de los modelos de TRI y es de capital importancia en las aplicaciones. La formulación matemática de las CCI es lo que distingue unos modelos de otros.

De manera análoga, se puede hablar de curva característica del test (CCT), que se define como la suma de las curvas características de los ítems que componen el test, o sea,

$$CCT = \sum_{i=1}^n P_i(\theta)$$

Nosotros utilizaremos el modelo para ítems de respuesta continua propuesto por Wang y Zeng (1998) el cual es una reparametrización del modelo de Samejima (1973) el cual es una forma límite del modelo de respuesta graduada. Este tipo de formato es muy poco usado en la práctica debido a que es poco usual obtener respuestas continuas en psicología y educación.

Un estudio profundo de este modelo se puede encontrar en Ferrando (2002) y en Wang y Zeng (1998).

En el modelo de Wang y Zeng (1998), la probabilidad de un sujeto i que tiene un rasgo θ obtenga un puntaje mayor o igual a x en el ítem j viene dado por:

$$P(X_{ij} \geq x / \theta) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^v e^{-\frac{t^2}{2}} dt$$

$$\text{con } v = a_j \left(\theta_i - b_j - \frac{1}{\alpha_j} \ln \frac{x_{ij}}{k_j - x_{ij}} \right)$$

El parámetro a representa la discriminación del ítem, b la dificultad, α es un parámetro de escala y k el máximo valor que puede tomar el puntaje.

Samejima (1973) obtiene los estimadores máximo versosimiles de las habilidades cuando los parámetros de los ítems son conocidos a partir de la ecuación:

$$MLE(\theta_i) = \frac{\sum_{j=1}^m \left[a_j^2 \left(b_j + \frac{z_{ij}}{\alpha_j} \right) \right]}{\sum_{j=1}^m \alpha_j^2}$$

El error estandar asintótico de la habilidad es:

$$SE(\theta_j) = \sqrt{\frac{1}{\sum_{j=1}^m \alpha_j^2}}$$

Para determinar el ajuste los θ son ordenados y asignados a k intervalos. Se calcula la media de los puntajes de los items \bar{X}_{mk} para cada item y cada intervalo. Se calculan la esperanza y la varianza condicional a la mediana de los θ para cada intervalo,

El *residuo estandarizado* para cada item en cada intervalo se obtiene como:

$$z_{mk} = \frac{\bar{X}_{mk} - E(X_m / \theta_k)}{\sqrt{\frac{V(X_m / \theta_k)}{N_k}}}$$

donde N_k es la cantidad de sujetos que caen en el intervalo k (Samejima, 1973).

Desde comienzos de este siglo, ya se hizo notar la importancia de aplicar la TRI al Rorschach (Meyer & Archer, 2001). Ello permitiría mejorar las propiedades psicométricas del test así como la generación de tests “personalizados” y “adaptativos”, entre otras muchas posibilidades que permite la TRI (Muñiz, 2010), tal como mencionamos anteriormente. A pesar de ello, no existen casi referencias respecto a la aplicación de la TRI al Rorschach. Realizando una búsqueda bibliográfica en el Metabusador TIMBÓ (www.timbo.org.uy) con las palabras clave “Rorschach” e “Item Response Theory” en el campo “Abstract” se halla solo un resultado (Hershberger & Segal, 2004), que sin embargo no refiere a la temática del presente artículo. Ampliando la búsqueda a otros medios “no académicos” como el buscador de Google, hallamos solamente tres antecedentes (Matsui, 1992a, 1992b, 1994). Los mismos aplicaron la TRI al Rorschach para analizar la calidad formal (Matsui, 1992b), las respuestas de movimiento humano y animal (Matsui, 1992a), y variables de localización, determinantes y contenidos (Matsui, 1994). El hecho de que estos trabajos estén publicados en japonés y que no estén indexados en las principales bases de datos ha dificultado su difusión. Entendiendo que se trata de una temática relevante y poco investigada, es que se lleva adelante el presente estudio.

Materiales y Método

El diseño metodológico utilizado en el presente estudio es de tipo descriptivo transeccional (Hernández, Fernández, & Baptista, 2010).

Participantes

Para el presente trabajo se re-codificaron y analizaron los datos tomados para dos estudios previos (Jiménez Gómez, 1990a; Jiménez Gómez, De Diego, & Sánchez, 1992). En los mismos, se realizó un muestreo por conveniencia a 300 mayores de 18 años residentes en Castilla y León (España) entre los años 1981 y 1984. De la muestra total, algunos casos no pudieron utilizarse para el presente estudio por estar incompletos o ilegibles, pudiendo re-codificarse un total de 224 casos. La Tabla 1 describe las principales características de la muestra en cuanto a sexo y edad.

Tabla 1- Descripción de la muestra (sexo y edad)

Sexo	N	Media	D.T.	Mínimo	Máximo
Femenino	112	33,009	11,3022	18,0	60,0
Masculino	112	39,196	12,6026	19,0	60,0
Total	224	36,103	12,3391	18,0	60,0

Instrumentos

A los participantes en el estudio se les realizó una entrevista semiestructurada, y posteriormente se les aplicaron dos tests psicológicos:

- Test de Rorschach (Rorschach, 1921), el cual fue administrado y codificado mediante el método de la escuela centroeuropea (Bohm, 1973).
- Las Láminas Proyectivas elaboradas por el Dr. Rodríguez Isidoro (Jiménez Gómez, 1990a). Las mismas son similares a las láminas del Rorschach, manteniendo los principios de simetría y de una baja estructuración estimular. Se administraron y codificaron mediante el sistema de la escuela centroeuropea, con las normas españolas (Jiménez Gómez, 1990a, 1990b).

Procedimiento

Los participantes fueron reclutados mediante un muestreo por conveniencia, todos ellos participaron voluntariamente de acuerdo a las normas y procedimientos éticos vigentes. Se realizaba una entrevista semiestructurada breve, y posteriormente se procedía a la aplicación de los tests anteriormente referidos. La codificación de los tests se realizó mediante la metodología referida anteriormente (Bohm, 1973; Jiménez Gómez, 1990b). Una vez codificados, para posibilitar el análisis de los mismos a través de la TRI, se recodificaron los mismos en dos etapas. En la primera, elaboraron escalas “Likert” para cada respuesta, en función del siguiente patrón:

A. Localización.

- a. Globales (con o sin espacio en blanco): 2
- b. Detalle grande (con o sin espacio en blanco): 1
- c. Detalle pequeño y detalle oligofrénico (con o sin espacio en blanco): 0

B. Determinantes:

a. Movimiento:

i. Movimiento humano

1. Presencia: 1
2. Ausencia: 0

ii. Movimiento animal

1. Presencia: 1
2. Ausencia: 0

iii. Movimiento inanimado:

1. Presencia: 1
2. Ausencia: 0

b. Color (en el sistema de Bohm no se distingue color cromático del acromático):

- i. Color puro: 3
- ii. Color-Forma: 2
- iii. Forma-Color: 1
- iv. Sin color: 0

- c. Claroscuro:
 - i. Claroscuro puro: 3
 - ii. Claroscuro-Forma: 2
 - iii. Forma-Claroscuro: 1
 - iv. Sin claroscuro: 0

- d. Textura:
 - i. Textura pura: 3
 - ii. Textura-Forma: 2
 - iii. Forma-Textura: 1
 - iv. Sin textura: 0

- e. Calidad formal:
 - i. Superior (+): 3
 - ii. Ordinaria (o): 2
 - iii. Ambigua (+/-): 1
 - iv. Mala (-): 0

C. Contenidos:

- a. Humanos
 - i. Cualquier contenido humano: 1
 - ii. Otros: 0
- b. Animales:
 - i. Cualquier contenido animal: 1
 - ii. Otros: 0
- c. Vestimenta, alimentos y sexual los mismos se agrupan debido a que algunos autores los consideran indicadores:
 - i. Contenidos Vestimenta, Alimento o Sexual: 1
 - ii. Otros: 0
- d. Otros:
 - i. Contenidos Humano, Animal, Vestimenta, Alimentos o sexual: 0
 - ii. Otros: 1

D. Frecuencia:

- a. Original: 2
- b. No especificado: 1

c. Vulgar: 0

Los contenidos Vestimenta, Alimento y Sexo se agruparon juntos, debido a que algunos autores consideran que los tres son indicadores de de “oralidad”, o de “descarga de impulsos poco modulada”(Bohm, 1973, 1979). Posteriormente, se agrupan el resto de contenidos bajo el supuesto de que responden a otro tipo de intereses de tipo intelectualizados (Bohm, 1973, 1979; Exner, 1969).

Para poder aplicar el modelo de análisis propuesto, se procedió posteriormente a seleccionar un único valor para cada lámina. Dicho valor es entonces el promedio de los valores obtenidos en todas las respuestas para cada lámina, en cada una de las variables y según los valores anteriormente referidos. Estos promedios por variable y por lámina son los que se analizarán.

Una vez codificados los datos según el método mencionado, se aplicó el modelo de TRI denominado “Modelo de Respuesta Continua” (Wang & Zeng, 1998), presentado anteriormente. Se utilizó el paquete EstCRM del software estadístico “R”.

Para determinar el grado de ajuste al modelo, se calcularon los Residuos Estandarizados, definidos anteriormente. Se considera un ajuste “muy bueno” si los mismos oscilan entre $\pm 1,96$, y un ajuste “adecuado” si los mismos oscilan entre $\pm 3,00$ (Samejima, 1973).

Resultados

Se realizaron los análisis en las distintas variables de los a (discriminación del ítem), b (dificultad del ítem), y $Alpha$ (parámetro del ítem), en cada lámina. Se utiliza numeración romana para las láminas del Rorschach, tal como es habitual, y las numeraciones +I, +II y +III para las 3 láminas proyectivas de Rodrigues Isidoro (Jiménez Gómez, 1984, 1990a).

En el caso de algunas variables, dichos parámetros no pudieron ser estimados, debido a la muy baja frecuencia de respuestas de dicha categoría. Ello ocurrió en las variables: Movimiento Inanimado, Clarooscuro y Textura.

La Tabla 2 muestra los estadísticos descritos, en cada lámina, en cuanto a Localización. Como puede observarse, todos los estadísticos respecto al nivel

de dificultad (b) se agrupan en el rango negativo, con escasa variabilidad entre ellos. Ello sesga de manera considerable la fiabilidad de la escala a un determinado sub-grupo de población con un nivel en la variable similar a dichos niveles de dificultad, con fiabilidades menores en el resto de la misma.

Tabla 2-Estadísticos a , b y Alpha en: Localización

Estadístico	a	b	Alpha
I	0.434138	-3.73344	0.917453
II	0.195846	-5.25624	0.450281
III	0.447836	-3.25001	0.815472
IV	0.326057	-4.40751	0.830441
V	0.415932	-5.72801	0.776764
VI	0.497744	-2.5638	1.177672
VII	0.445969	-1.87894	1.009268
VIII	0.570299	-2.01102	1.059165
IX	0.380019	-1.5348	0.943996
X	0.335469	-1.97665	0.554625
+I	0.864159	-2.03217	1.624805
+II	1.1622	-1.33473	1.815575
+III	0.627706	-1.43738	1.270922

La Tabla 3 los estadísticos de las respuestas de movimiento sobre las que se pudo realizar dichos análisis (no se pudo analizar el movimiento inanimado). En este caso ambas variables presentan una gran variabilidad en cuanto a los estadísticos b . Si bien, como dijimos anteriormente, la fiabilidad depende de cada sujeto y su nivel en cada variable, a grandes rasgos podríamos decir que

es un indicador que podría favorecer una mayor fiabilidad de la escala en estas variables.

Tabla 3- Estadísticos a, b y Alpha en: Movimiento Humano (M), y Movimiento Animal (FM)

Estadístico	Movimiento Humano			Movimiento Animal		
	a	b	Alpha	a	b	Alpha
I	0,424766	7,263029	0,624682	0,304995	31,59272	0,105336
II	0,346496	5,60584	0,670807	0,171327	28,9829	0,109398
III	0,507117	2,674777	0,862707	0,194436	86,23643	0,03899
IV	0,230411	12,18252	0,370079	0,156965	47,5103	0,069485
V	0,306821	16,76513	0,298079	0,223742	19,23029	0,16241
VI	0,140155	26,89333	0,18048	0,059051	299,1416	0,011275
VII	0,428047	5,48249	0,750944	0,720179	9,893703	0,325567
VIII	0,200746	19,31352	0,250071	0,27799	21,04205	0,152875
IX	0,301784	9,639866	0,467367	0,000976	12710,54	0,000264
X	0,539365	6,772115	0,664797	0,339785	20,88131	0,153754
+I	0,361758	4,267883	0,720162	0,433949	21,1359	0,156835
+II	0,508045	2,969302	0,947076	0,130713	44,64834	0,072784
+III	0,367031	6,759081	0,615351	0,347589	16,40802	0,194937

La Tabla 4 muestra los resultados del determinante Color y Calidad Formal. En cuanto a la primera variable también hay una gran dispersión de los niveles de dificultad de cada variable, mientras que en el caso de la calidad formal los mismos se encuentran muy agrupados dentro del rango 0 a -3. Tal como mencionamos anteriormente, una mayor dispersión de los *b* es, en principio, un

indicador de una mayor fiabilidad.

Tabla 4- Estadísticos a, b y Alpha en: Color (C), y Calidad Formal (FQ)

Estadístico	Color			Calidad Formal		
	a	b	alpha	a	b	Alpha
I	0,185542	31,11059	0,175678	0,079404	-1,92043	0,268976
II	0,470272	3,200455	1,047728	0,155733	-0,1373	0,415888
III	0,342101	5,451874	0,711296	0,297208	-1,37311	0,59068
IV	0,047807	155,8091	0,035683	0,349515	-0,51642	1,192566
V	0,264985	24,71811	0,223476	0,310612	-3,05062	0,928446
VI	0,188535	39,71806	0,140375	0,178844	-2,82212	0,635302
VII	0,187603	35,68145	0,154855	0,342945	-0,67566	0,998926
VIII	0,561276	2,910787	1,13487	0,415058	-1,1341	1,01997
IX	0,476201	2,237877	1,116337	0,257954	1,29533	0,73161
X	0,526238	2,733148	0,933627	0,419102	0,044616	0,809962
+I	0,427858	5,189234	0,810875	0,30877	-1,78203	0,738391
+II	0,445639	4,161441	0,897102	0,570414	-1,26181	1,132412
+III	0,377472	4,380557	0,790247	0,567783	-0,67055	1,39764

La Tabla 5 analiza las respuestas de contenido Humano y Animal (las más frecuentes en el Rorschach), mientras que la Tabla 6 presenta el resto de contenidos. En el caso de los contenidos humano, animal, y V, A, Sex, los *b* están relativamente dispersos, mientras que en el caso de “Otros” los mismos están más agrupados.

Tabla 5- Estadísticos a, b y Alpha en: Contenidos Humanos (Humano), y Animal (Animal)

Estadístico	Contenido Humano			Contenido Animal		
	a	b	alpha	a	b	alpha
I	0,450582	5,480775	0,778148	0,228531	5,501933	0,348103
II	0,426283	6,086356	0,728157	0,270924	4,837594	0,482448
III	0,215549	7,149154	0,289762	0,335101	5,599884	0,618996
IV	0,455756	5,499498	0,797807	0,208984	5,70513	0,40553
V	0,414171	9,397648	0,539894	0,25298	4,798895	0,217679
VI	0,374844	8,678842	0,560118	0,236094	4,628585	0,41218
VII	0,432426	4,197429	0,836853	0,596402	3,143765	1,08685
VIII	0,423792	9,551535	0,533286	0,000965	1539,19	0,001062
IX	0,564452	3,763505	1,026209	0,350636	6,608851	0,642148
X	0,710231	4,264833	1,008949	0,309861	4,441969	0,483669
+I	0,149572	8,356009	0,273435	0,280773	6,199011	0,560459
+II	0,102867	11,78138	0,151041	0,16946	14,50903	0,305222
+III	0,391584	4,670346	0,737972	0,467239	3,794754	0,870602

Tabla 6- Estadísticos a, b y Alpha en: Contenidos Vestimenta, Alimento y Sexo (V, A, Sex) y Otros Contenidos (Otros)

	Vestimenta, Alimento y Sexo			Otros contenidos		
	a	b	alpha	a	b	alpha
I	0.224704	32.01222	0.172296	0.221082	8.517721	0.436638
II	0.364385	14.68917	0.364513	0.308101	5.644639	0.606471
III	0.211579	18.65275	0.273925	0.327083	5.693408	0.612776

IV	0.165248	25.65083	0.203357	0.243094	7.618161	0.495637
V	0.053312	235.6991	0.023931	0.411318	9.509894	0.537501
VI	0.92673	7.268188	0.735038	0.551123	4.072967	0.983805
VII	0.815257	8.867024	0.61022	0.389621	4.318074	0.780289
VIII	0.000492	11246.82	0.00048	0.409929	4.085254	0.754514
IX	0.163369	29.49323	0.180122	0.232073	5.478996	0.480412
X	0.19294	26.75755	0.199144	0.304415	5.513154	0.543943
+I	0.598423	17.45662	0.320171	0.441905	5.545655	0.772942
+II	0.080023	35.98277	0.129665	0.348818	5.75992	0.63937
+III	0.449391	6.257258	0.722802	0.341716	5.553496	0.655566

Por último, la Tabla 7 presenta los estadísticos relativos a la variable Frecuencia. La dispersión de b es también relativamente baja.

Tabla 7- Estadísticos a , b y α en: Frecuencia

	a	b	alpha
I	0,235561	3,306842	0,45018
II	0,251751	2,827007	0,511226
III	0,352914	2,872215	0,691752
IV	0,485377	2,448646	0,897444
V	0,415761	3,105152	0,958736
VI	0,458283	2,017618	1,022469
VII	0,348208	2,198434	0,759186
VIII	0,47655	2,065836	0,960891

IX	0,284498	1,212394	0,459593
X	0,436785	1,658751	0,572341
+I	0,46058	2,632253	1,044671
+II	0,306616	4,264476	0,710359
+III	0,439317	1,799195	0,90381

Para terminar, y haciendo referencia al principal objetivo del presente estudio, la Tabla 8 muestra los Residuos Estandarizados (R.E.) por cada lámina y variable estudiada. Los R.E. son utilizados como estadístico que nos permite calcular el grado de ajuste de los datos al modelo, bajo los supuestos anteriormente mencionados del Modelo de Respuesta Continua de la Teoría de Respuesta al Ítem (Wang & Zeng, 1998). Si interpretamos los resultados de la Tabla 8 en función de una descripción del grado de ajuste al modelo, tal como puede observarse, de los 130 R.E. estudiados, 64 están dentro del rango considerado como “adecuado”, mientras que 42 de las mismas se encuentran dentro del rango considerado como “muy bueno”. Agrupando las variables en función del grado de ajuste al modelo (interpretación cualitativa de los datos) podríamos definir 3 categorías:

- Variables que claramente no ajustaron adecuadamente al modelo:
 - Localización
 - Calidad Formal
- Variables que mostraron un ajuste al modelo “adecuado”:
 - Movimiento Humano
 - Movimiento Animal
 - Contenidos Vestimenta, Alimento y Sexo
 - Frecuencia
- Variables que mostraron un muy buen ajuste al modelo:
 - Color

- Contenido Humano
- Contenido Animal
- Otros contenidos

Por otra parte, se exponen también las fiabilidades de dichas escalas, utilizando el estadístico α de Cronbach. Las mismas oscilaron entre 0,73 y 0,41, valores que podrían considerarse bajos.

Tabla 8- Residuos Estandarizados promedios por Variable y Lámina, y fiabilidades (α de Cronbach)

	Loc	M	FM	Color	FQ	Humano	Animal	V, A, Sex	Otros	Frecuencia
I	44,311	2,177	2,465	1,605	8,769	2,144	1,226	2,381	1,732	3,351
II	29,785	2,536	3,157	1,605	8,280	2,142	1,645	3,191	1,589	3,811
III	50,568	4,155	2,476	1,561	12,736	1,093	1,582	3,982	1,543	2,566
IV	29,722	2,733	2,080	1,426	9,532	1,857	2,054	2,853	1,816	2,562
V	74,018	2,378	2,649	1,970	21,543	2,364	-0,423	1,399	2,907	2,595
VI	39,261	2,125	1,584	1,736	12,013	2,761	2,155	-	2,094	3,779
VII	24,004	2,169	5,639	1,522	10,720	1,591	2,379	3,496	1,814	4,076
VIII	32,199	2,249	2,834	1,670	12,730	2,613	0,958	2,555	1,695	3,417
IX	18,153	2,627	2,116	2,450	5,950	1,904	1,947	2,530	2,133	5,921
X	24,702	3,778	3,987	1,740	8,885	2,862	1,416	2,551	1,394	2,403
+I	64,750	2,607	3,301	1,714	16,015	1,622	1,610	2,218	2,091	2,622
+II	57,606	3,546	2,360	1,732	21,142	1,292	1,963	4,406	1,836	2,217
+III	27,730	2,704	2,934	1,756	14,714	1,635	1,718	4,412	1,699	4,103
α Cr.	0,73	0,58	0,41	0,63	0,57	0,53	0,49	0,52	0,53	0,63

Si bien se observan diversos matices y variantes, podría decirse que respecto a las distintas láminas, todas presentan similares ajustes al modelo en general, con variaciones para algunas variables específicas en cada caso. Esto se dio tanto para las láminas del Rorschach como para las Láminas Proyectivas.

Discusión

Para un estudio como el presente, con tan pocos antecedentes, resulta particularmente interesante el análisis de los resultados una vez efectuados los análisis correspondientes. Francamente los resultados, en cuanto al ajuste al modelo, fueron mejores a los esperables. Varias limitaciones hacían pensar previamente que el ajuste podría presentar dificultades. En primer lugar, la casi ausencia de antecedentes en los que apoyarse, pocos estudios previos (Matsui, 1992a, 1992b, 1994), a cuyos resultados además hubo dificultades en el acceso por barreras lingüísticas.

Por otra parte, el hecho de que el sistema de aplicación del Rorschach que se utilizó en la recogida de datos, fuera el de Bohm, no permitía disponer de todas las mejoras que el Sistema Comprensivo (Exner, 1969) otorgó al Rorschach, particularmente en cuanto a sus propiedades psicométricas y definición más precisa de constructos y criterios. Creemos que ello puede ser un punto clave para comprender el escaso ajuste al modelo que presentaron las variables Localización y Calidad Formal. En el Sistema Comprensivo hay una redefinición de la manera en que se codifican ambas variables.

En el caso de la Localización, en lugar de usarse un criterio de “tamaño de la zona de la mancha”, se utiliza el criterio de “frecuencia de la zona de la mancha” (Exner, 2007). Evidentemente se está codificando la variable de manera muy diferente, y se están evaluando aspectos cognitivos también distintos. En su momento Exner optó por esta forma de codificación porque le ofrecía un mayor apoyo empírico y una mayor fiabilidad (Exner, 1969).

En el caso de la calidad formal, también Exner introduce variantes que

permiten mejorar la exactitud en la codificación. Un extensísimo repertorio de respuestas más frecuentes orientan al codificador en la asignación de la correspondiente calidad formal, con un sistema de pre-asignación en modalidad inter-jueces de la FQ que corresponde a cada respuesta en cada zona de la mancha (Exner, 2007). Ello reduce muchísimo el margen de error, y las posibles arbitrariedades en la asignación “a ojo” por el aplicador de la prueba.

Respecto a los contenidos, la agrupación por la que optamos para el presente trabajo, presentó fortalezas y debilidades. Si bien los datos se ajustaron de manera bastante razonable al modelo, entendemos que habrá que seguir trabajando en la forma de re-codificación de los contenidos, en función de los hallazgos del presente estudio. La agrupación de todas las respuestas de contenido Humano juntas, de todas las respuestas de tipo Animal juntas, y las dos categorías de “V, A, Sex” y “Otros” deberán ser reevaluadas en futuros estudios.

A pesar de que el ajuste al modelo se puede considerar razonablemente “adecuado” si se tienen en cuenta los “Residuos Estandarizados Promedio”, es necesario observar otros indicadores que muestran dificultades en el ajuste. Se encontraron en varias escalas ítems con valores de dificultad alejados del rango de respuesta. Ello implica que los mismos no aportan información “alta” en dicho rango, y no aportan a la estimación del rasgo considerado. Consideramos que se debe revisar la puntuación en el continuo de estos ítems o evaluar la posibilidad de no considerarlos para la estimación del rasgo. El alejamiento de los ítems respecto a los rangos de respuesta es uno de los elementos fundamentales para las estimaciones acerca de la fiabilidad a través de la TRI. Tal como ocurre con cualquier otra escala, algunos ítems no ajustan y es necesario sustituirlos por otros. Ello no ha ocurrido con el test de Rorschach desde 1921. Este trabajo pretende justamente aportar para la construcción de criterios o estándares que permitan seleccionar, eliminar, crear, nuevos “ítems”, láminas, variables, etc.

También es destacable el funcionamiento de las Láminas Proyectivas, desconocidas en nuestro medio. Las mismas cuentan con varias publicaciones

y apoyo empírico, y se proponen como alternativa para “compensar” algunas deficiencias del Rorschach (Jiménez Gómez, 1990a). Ello sugiere la posibilidad de creación y utilización en futuras investigaciones de otras láminas, que puedan ampliar el repertorio sobre el cual elegir a la hora de diseñar futuros (posibles) tests adaptativos. Durante casi 100 años se han utilizado las mismas láminas, las mismas fueron seleccionadas en su momento bajo unos criterios y con unos datos empíricos muy diferentes a los que disponemos hoy (Rorschach, 1921). Nada podría hacernos suponer que esas 10 láminas poseen alguna propiedad “mágica” o “especial” que las haga distintas de otras que puedan crearse en el futuro. Ello abre la posibilidad de tests que se adapten a determinados requerimientos específicos, para evaluar con mayor precisión determinados puntos de corte, determinados perfiles de población, adaptativos, versiones largas y versiones de screening, etc.

Las fiabilidades obtenidas mediante el método utilizado en el presente estudio no son demasiado altas. Es necesario seguir trabajando en la detección de las causas de ello. En el estudio del Rorschach es habitual utilizar como indicador de fiabilidad criterios como el “inter-jueces” (Exner, 1969, 2007; Meyer et al., 2002). Este estudio utiliza otras formas de acercamiento a dicha relevante propiedad psicométrica, más usuales en otros instrumentos. Podría decirse que este estudio representó un intento de “normalizar” un test claramente atípico como es el Rorschach, y acercarlo a parámetros compartidos por la mayoría de tests y escalas psicológicas.

El Rorschach es además un test que implica tareas principalmente visoperceptivas, y no tanto verbales (Acklin & Wu-Holt, 1996; Selma, en prensa). Por ello, creemos que es un test con potencial de ser menos influenciado por variables lingüísticas y culturales que otros, por ejemplo cuestionarios. Ello permitiría, en conjunción con las posibilidades que abre el TRI, el desarrollo de estudios transculturales con mucha mayor facilidad y potencial (Muñiz, 1997, 2010). A ello además se agrega la menor influencia que ejercen sobre el test variables como la deseabilidad social y las distorsiones en el autoconocimiento, que clásicamente son un problema de los cuestionarios, tests de amplia difusión a nivel clínico (Meyer & Archer, 2001).

En definitiva, la aplicación de la TRI al Rorschach, si es posible, es decir, si se logra que las variables estudiadas se “ajusten” adecuadamente al modelo, permitiría abrir una interesante y promisoría línea de investigación. Como acercamiento a la aplicación de la TRI al Rorschach, creemos que los resultados son alentadores, y será necesario continuar investigando en esta línea para poder llegar al desarrollo futuro de nuevas formas de aplicar, codificar e interpretar el test.

Agradecimientos

El presente artículo es parte de la elaboración de una Tesis Doctoral bajo el formato “Compendio de publicaciones”. No hubiera sido posible sin la colaboración de mis Directores de Tesis, el Dr. Fernando Jiménez y la Dra. Guadalupe Sánchez de la Universidad de Salamanca, y de la Comisión Sectorial de Investigación Científica de la Universidad de la República. Agradezco especialmente al Dr. Borja Jordán de Urríes por su invaluable ayuda.

Tesis Doctoral: *Fiabilidad y Validez del Test de Rorschach: aportes de la Teoría de Respuesta al Ítem y de los Modelos Psicobiológicos de la Personalidad de Cloninger y Zuckerman*. Institución financiadora: *Universidad de la República*. Período de realización: 2012-2015.

Referencias

- Abad, F., Olea, J., Ponsoda, V., & García, C. (2011). *Medición en Ciencias Sociales y de la Salud*. Madrid: Síntesis.
- Acklin, M. W., & Wu-Holt, P. (1996). Contributions of cognitive science to the Rorschach Technique: cognitive and neuropsychological correlates of the response process. *Journal of Personality Assessment*, 67(1), 169–178.
http://doi.org/10.1207/s15327752jpa6701_13
- Bohm, E. (1973). *Manual del psicodiagnóstico de Rorschach*. Morata.
- Bohm, E. (1979). *El test de Rorschach*. Morata.

- Exner, J. E. (1969). *The Rorschach Systems*. New York: Grune & Stratton.
- Exner, J. E. (2007). *Manual de codificación del Rorschach para el Sistema Comprehensivo* (3a ed.). Madrid: Psimática.
- Ferrando, P. J. (2002). Theoretical and Empirical Comparison between Two Models for Continuous Item Responses. *Multivariate Behavioral Research*, 37(4), 521–542.
- Ganellen, R. J. (2001). Weighing Evidence for the Rorschach's Validity: A Response to Wood et al. (1999). *Journal of Personality Assessment*, 77(1), 1–15.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. SAGE Publications.
- Hernández, R., Fernández, C., & Baptista, M. P. (2010). *Metodología de la investigación* (5th ed.). México: McGraw Hill.
- Hershberger, S. L., & Segal, N. L. (2004). The Cognitive, Behavioral, and Personality Profiles of a Male Monozygotic Triplet Set Discordant for Sexual Orientation. *Archives of Sexual Behavior*, 33(5), 497.
- Hibbard, S. (2003). A Critique of Lilienfeld et al.'s (2000) "The Scientific Status of Projective Techniques." *Journal of Personality Assessment*, 80(3), 260–271.
- Hiller, J. B., Rosenthal, R., Bornstein, R. F., Berry, D. T. R., & Brunell-Neuleib, S. (1999). A comparative meta-analysis of Rorschach and MMPI validity. *Psychological Assessment*, 11(3), 278–296. <http://doi.org/10.1037/1040-3590.11.3.278>
- Hunsley, J., & Michael, J. (1999). The clinical utility of the Rorschach: Unfulfilled promises and an uncertain future. *Psychological Assessment*, 11(3), 266–277. <http://doi.org/10.1037/1040-3590.11.3.266>
- Jiménez Gómez, F. (1984). Láminas Proyectivas: aportación complementaria al Psicodiagnóstico de Rorschach. Presented at the XI Congreso Internacional del Rorschach y Métodos Proyectivos, Barcelona.

- Jiménez Gómez, F. (1990a). Complementariedad de las láminas Rorschach y proyectivas: una demostración empírica. *Revista de Psicología*, *8*(2), 155–179.
- Jiménez Gómez, F. (1990b). *Introducción al psicodiagnóstico de Rorschach y Láminas Proyectivas*. Salamanca: Amarú Ediciones. Retrieved from <http://www.agapea.com/libros/Introduccion-al-psicodiagnostico-de-Rorschach-y-laminas-proyectivas-9788486368258-i.htm>
- Jiménez Gómez, F., De Diego, R., & Sánchez, G. (1992). El movimiento humano en las láminas proyectivas: una aportación a la III del test de Rorschach. *Revista de Psicología*, *9*(1-2), 65–81.
- Louitt, C. M., & Browne, C. G. (1947). Psychometric instruments in psychological clinics. *Journal of Consulting Psychology*, *11*, 49–54.
- Lumsden, J. (1961). The construction of unidimensional tests. *Psychological Bulletin*, *58*(2), 122–131. <http://doi.org/10.1037/h0048679>
- Márquez Sánchez, M. O. (1986). Estudio bibliométrico sobre el test de Rorschach. *Revista de Historia de La Psicología*, *7*(3), 71–82.
- Matsui, H. (1992a). Application of item response theory for the inkblot test. *Japanese Journal of Educational Psychology*, *40*, 29–36.
- Matsui, H. (1992b). Scaling Technique on Form Definiteness of the Inkblot Test. *Bulletin of the Faculty of Education, University of Tokyo*, *31*, 173–182.
- Matsui H. (1994). Optimum Split-Half Reliabilities for the Rorschach Test. *Journal of Projective Technique*, *24*, 171–181.
- McDonald, R. P. (1981). The dimensionality of tests and items. *British Journal of Mathematical and Statistical Psychology*, *34*(1), 100–117. <http://doi.org/10.1111/j.2044-8317.1981.tb00621.x>
- Meyer, G. J., & Archer, R. P. (2001). The hard science of Rorschach research: what do we know and where do we go? *Psychological Assessment*, *13*(4), 486–502.
- Meyer, G. J., Hilsenroth, M. J., Baxter, D., Exner, J. E., Fowler, J. C., Piers, C.

- C., & Resnick, J. (2002). An examination of interrater reliability for scoring the Rorschach Comprehensive System in eight data sets. *Journal of Personality Assessment, 78*(2), 219–274.
http://doi.org/10.1207/S15327752JPA7802_03
- Meyer, G. J., Riethmiller, R. J., Brooks, R. D., Benoit, W. A., & Handler, L. (2000). A Replication of Rorschach and MMPI-2 Convergent Validity. *Journal of Personality Assessment, 74*(2), 175–215.
- Muñiz, J. (1997). *Introducción a la Teoría de Respuesta a los Ítems*. Madrid: Pirámide.
- Muñiz, J. (2003). *Teoría clásica de los tests*. Pirámide.
- Muñiz, J. (2010). Test Theories: Classical theory and item response theory. *Papeles Del Psicólogo, 31*(1), 57–66.
- Park, K. N. (2009, May 23). *A Review of Literature Regarding Scientific Controversies Surrounding the Psychometric Properties of the Rorschach Inkblot Test* (Tesis Doctoral). Biola University. Retrieved from <http://www.eric.ed.gov/ERICWebPortal/contentdelivery/servlet/ERICServlet?accno=ED506766>
- Rorschach, H. (1921). *Psychodiagnostik*. París: PUF.
- Samejima, F. (1973). Homogeneous case of the continuous response model. *Psychometrika, 38*(2), 203–219. <http://doi.org/10.1007/BF02291114>
- Selma, H. (en prensa). Rorschach y psicobiología de la personalidad. *Universitas Psychologica, 15*(1).
- Society for Personality Assessment. (2005). The status of the Rorschach in clinical and forensic practice: an official statement by the Board of Trustees of the Society for Personality Assessment. *Journal of Personality Assessment, 85*(2), 219–237.
http://doi.org/10.1207/s15327752jpa8502_16
- Sundberg, N. D. (1961). The practice of psychological testing in clinical services in United States. *American Psychologist, 16*, 79–83.

- Wang, T., & Zeng, L. (1998). Item Parameter Estimation for a Continuous Response Model Using an EM Algorithm. *Applied Psychological Measurement, 22*(4), 333–44.
- Wood, J. M., Nezworski, M. T., & Stejskal, W. J. (1996). The Comprehensive System for the Rorschach: A Critical Examination. *Psychological Science, 7*(1), 3–10. <http://doi.org/10.1111/j.1467-9280.1996.tb00658.x>
- Wood, J. M., Nezworski, M. T., & Stejskal, W. J. (1997). The reliability of the Comprehensive System for the Rorschach: A comment on Meyer (1997). *Psychological Assessment, 9*(4), 490–494. <http://doi.org/10.1037/1040-3590.9.4.490>
- Wood, J. M., Nezworski, M. T., Stejskal, W. J., Garven, S., & West, S. G. (1999). Methodological issues in evaluating Rorschach validity: a comment on Burns and Viglione (1996), Weiner (1996), and Ganellen (1996). *Assessment, 6*(2), 115–129.