

Modelización de ítems de matrices figurales y pautas específicas propuestas para su construcción

G. Diego Blum*, Gabriela S. Lozzia, Facundo J.P. Abal, y Horacio F. Attorresi

Instituto de Investigaciones de la Facultad de Psicología de la Universidad de Buenos Aires

Resumen: Este artículo da a conocer los resultados de la modelización de un Test de Analogías Figurales (TAF) utilizando los aportes conjuntos de la Teoría Clásica de Tests (TCT) y de la Teoría de Respuesta al Ítem (TRI), así como la posterior interpretación de las desventajas y posibles soluciones del diseño de dicho test. Se desarrolla el marco conceptual que dio origen al armado del TAF, así como los estudios realizados con esta prueba al administrarla a dos muestras. Una de las muestras corresponde a un estudio piloto con estudiantes de Psicología, mientras que la otra muestra consistió en estudiantes de carreras artísticas y de diseño. Los resultados principales revelan indicadores favorables de unidimensionalidad y confiabilidad, así como parámetros aceptables de discriminación (*a*), dificultad (*b*) y pseudoazar (*c*). Se discuten las desventajas del diseño de ítems con tres reglas y nuevas especificaciones en cuanto al armado de reactivos matriciales de 2x2.

Palabras clave: Matrices; figuras; ítems; analogía; reglas.

Title: Modelling figural matrix items and specification of guidelines for their construction.

Abstract: This manuscript publicizes the results that concern the modeling of a Test of Figural Analogies (TFA) using Classical Test Theory (CTT) and Item Response Theory (IRT) on a joint basis, as well as the interpretation of the disadvantages and possible solutions of the test design. The theoretical background for the TFA construction is outlined, and studies accomplished after its administration to two samples is explained. One of these samples corresponds to a pilot study with Psychology students, while the other sample consisted of Art and Design college students. Main results indicate good unidimensionality and reliability, as well as acceptable discrimination (*a*), difficulty (*b*) and guessing value (*c*) parameters. Design disadvantages of three-rule based items are discussed, and new specifications for the 2x2 matrix item construction are given.

Key words: Matrices; figures; items; analogy; rules.

Introducción

Las matrices de figuras suelen utilizarse para medir componentes de la Inteligencia General (Freund, Hofer, & Holling, 2008), siendo uno de los más importantes el Razonamiento Analógico (RA. Sternberg, 1987). En particular, las matrices de 2x2 pueden emplearse para construir analogías de tipo A:B::C:D (Blum, Abal, Lozzia, Picón Janeiro y Attorresi, 2011). Es posible encontrar éste y otros modelos matriciales en el Test de Matrices Progresivas de Raven (Raven, Raven, & Court, 1991), en el Test de Cattell y Cattell (1973), en el Test de Inteligencia No-Verbal (Test of Non-Verbal Intelligence, TONI) de Brown, Sherbenou y Johnsen (2000), en los trabajos de Wolf Nelson y Gillespie (1991), en los trabajos de Susan Embretson (Embretson & Reise, 2000), en el subtest de Razonamiento con Matrices del WAIS-III (Wechsler, 1997), entre otros.

Numerosos estudios de calidad psicométrica se han realizado sobre tests como los que se mencionan, tanto desde la Teoría Clásica de Tests (TCT) como desde la más reciente Teoría de Respuesta al Ítem (TRI). En cuanto a esta última, Raven, Raven y Court (1991) emplearon el examen visual de las Curvas Características de los Ítems (CCI) para proporcionar información sobre la naturaleza de las aptitudes evaluadas y sobre el posible perfeccionamiento del test. Embretson (Embretson & Reise, 2000) presentó los resultados de la modelización de sus 30 ítems de Razonamiento Abstracto (Abstract Reasoning Test, ART) tanto desde el Modelo de Rasch como desde los de dos y tres parámetros (ML2P y ML3P), entre otros modelos aplicados. También se realizaron análisis del Funcionamiento Diferencial de los Ítems

(FDI) en pruebas de analogías y/o matrices. Algunas referencias pueden encontrarse en Abad, Colom, Rebollo y Escorial (2004), Bandeira Andriola (2000) y Maller (2000).

El RA es un razonamiento no deductivo y, como tal, no asegura una conclusión verdadera cuando todas sus premisas son verdaderas. Esto quiere decir que es un razonamiento que no está basado en una conclusión necesaria, sino más bien probable. Por este motivo, se trata de un razonamiento ampliatorio, que permite razonar sobre lo posible y no sobre lo deductivamente válido. A diferencia de otros razonamientos no deductivos, como la inducción y la abducción, en el RA se establecen inferencias hacia casos particulares partiendo de premisas sobre otros casos particulares, de manera que nunca se parte de una generalidad ni tampoco se arriba a una generalidad de casos. Esto quiere decir que el RA *va de lo particular a lo particular*.

La inferencia se realiza sobre un caso nuevo, poco conocido. El RA consiste en atribuirle a dicho caso propiedades que pertenecen a otros casos, los cuales son mejor conocidos que el anterior y de los que se establecen las premisas. En palabras de expertos como Gick y Holyoak (1980), Holyoak y Koh (1987) y Gentner (1983), el dominio nuevo o caso poco conocido se conoce como Análogo-Meta (*target analog*, TA) y el/los dominio(s) fuente o caso(s) mejor conocido(s) como Análogo-Fuente (*source analog*, SA). Pero esta relación sólo es posible si tanto el TA como el SA comparten previamente ciertas propiedades comunes que los hacen similares uno respecto del otro. Esto queda muy claro en la experiencia relatada por Gick y Holyoak (1980), sobre la investigación General - Radiación. En ella, para que pueda extrapolarse la solución al problema del general hacia el problema de la radiación, tiene que existir antes una serie de correspondencias entre los roles que asumen los componentes de ambos problemas. Por ejemplo, en los dos relatos encon-

* Dirección para correspondencia [Correspondence address]:

G. Diego Blum. Anchorena 1.169 3°B (1425), CABA Argentina.

E-mail: blumworx@gmail.com

tramos ciertos recursos para resolver el problema y también ciertas restricciones que impiden dicha resolución.

Por ello, la forma lógica de este razonamiento adquiere la siguiente expresión: si $P(X,Y)$ y $Q(X)$, entonces $Q(Y)$. Esto quiere decir que, como primera instancia, los dominios X e Y poseen en común la propiedad P , la cual los hace similares. Ya a partir de esta primera premisa, tales dominios se convierten en análogos uno del otro. Esto se conoce también como la estructura representacional compartida (Cubillo y González Labra, 1998), es decir, el conjunto de relaciones y roles que tienen en común ambos dominios. Pero además, de uno de esos dominios se conoce otra propiedad, que es Q , la cual lo convierte en un dominio mejor conocido. Por ello, dicho dominio pasa a ser el SA. Luego, se infiere que es probable que el otro dominio también posea la misma propiedad, lo cual convierte a este último en el TA.

Generalmente, los dominios son relaciones entre elementos y las analogías, por consiguiente, suelen concebirse como comparaciones entre relaciones, es decir, entre dominios (Sternberg, 1977). Esto se conoce mejor como analogías proporcionales del estilo $A:B::C:D$ (A es a B como C es a D), que son las analogías históricamente más estudiadas. Cuando un problema se basa en buscar el elemento faltante D de la analogía (*i.e.*, $A:B::C:?$), entonces $C:D$ se convierte en el TA y $A:B$ se convierte en el SA. El enfoque sintáctico de Gentner (1983) plantea que las analogías dependen de la proyección de los predicados sobre las relaciones entre los elementos de cada dominio, y no de la proyección de los predicados sobre dichos elementos. Es decir, para que una analogía se conforme como tal, debe proyectarse la estructura relacional de segundo orden que vincula a los dominios y no sólo las relaciones de primer orden entre sus elementos. En otros términos, no sólo debe compararse A con C o B con D para determinar la analogía, sino que más importante aún es la extrapolación de la relación implícita en el par $A:B$ hacia el otro par $C:D$. Bajo esta postura, una analogía determinada sólo por las relaciones de primer orden no es verdadera (De La Fuente Arnanz y Minervino, 2004). Sin embargo, desde el enfoque pragmático de Holyoak (*e.g.*, Holyoak & Thagard, 1989), las relaciones de primer orden también conforman analogías y los factores contextuales son más decisivos para el desarrollo de un RA que los estructurales. A lo largo del presente manuscrito se mantendrá la postura de Gentner (1983), la cual también fue enunciada en un trabajo anterior de los autores (Blum, Abal, Lozzia et al., 2011).

A su vez, dependiendo de sobre qué elementos se realizan estas relaciones y comparaciones, podemos concebir distintos tipos de analogías, por ejemplo las verbales, las pictóricas y las figurales. De estas últimas se han desarrollado estudios profusos en el campo de la psicometría y, como consecuencia, muchos tests de matrices de figuras conocidos miden el RA desde un enfoque proporcional. En la matriz, el individuo debe identificar los elementos relacionados del SA y los del TA, que suelen ser pares de figuras, y luego extrapolar mentalmente las relaciones mejor conocidas del SA para completar el TA. Debajo o a un costado de la matriz

aparecen opciones cerradas de respuesta. El individuo debe retener el TA que completó mentalmente y compararlo con cada una de las opciones presentadas, en orden de identificar la correcta o más cercana al elemento faltante del TA.

Desde un punto de vista cognitivo, la construcción de ítems basada en reglas puede jugar un rol importante en la evaluación del RA. Las reglas son operaciones cognitivas elementales que se requieren para resolver el ítem (Kubinger, 2008). Cada regla constituye “un determinado tipo de relación análoga” (Blum, Abal, Lozzia et al., 2011, p. 138), siendo ejemplos de reglas la adición (Arendasy, 2005; Freund, Hofer, & Holling, 2008), la sustracción (Arendasy, 2005), el desplazamiento espacial (*e.g.*, rotación) y distorsiones del tamaño y de la forma (Whitely & Schneider, 1981). Es posible crear reactivos con diferentes reglas o combinaciones de reglas (Kubinger, 2008; Zeuch, 2010). Blum, Abal, Lozzia et al. (2011) trabajaron con un conjunto de reglas para el diseño de ítems con matrices de 2×2 . Las reglas citadas por Blum, Abal, Lozzia et al. pueden agruparse en tres categorías: emplazamiento espacial (reglas de rotación, traslación y reflejo), distorsión (del tamaño y de la forma) y número (adición y sustracción). Los autores propusieron que el SA y el TA deberían ser, respectivamente, tanto $A:B$ y $C:D$ como $A:C$ y $B:D$, lo cual otorga dos caminos de resolución posibles. También desarrollaron maneras de evitar sesgos relacionados con el tipo de regla empleado, las opciones de contestación y las condiciones de administración de la prueba.

El objetivo del presente artículo es dar a conocer los resultados de la construcción y modelización de un Test de Analogías Figurales (TAF) a partir de la TCT y la TRI, ambas utilizadas en forma complementaria. Se trata de un artículo de corte metodológico y representa la continuación de un estudio descrito por Blum, Abal, Galibert y Attorresi (2011) y por Blum, Galibert, Abal, Lozzia y Attorresi (2011); dicho estudio previo se realizó sobre una muestra piloto de alumnos de la Facultad de Psicología de la Universidad de Buenos Aires (UBA). A diferencia del mismo, en el presente escrito se consideran además los datos de una segunda muestra de estudiantes provenientes de la Facultad de Arquitectura, Diseño y Urbanismo de la UBA y del Instituto Universitario Nacional del Arte (IUNA), así como su comparación con los datos de la muestra de Psicología. También se brindan sugerencias que se añaden a aquellas propuestas por Blum, Abal, Lozzia et al. (2011), para la sistematización y objetivación del proceso de construcción de ítems de matrices de figuras de 2×2 . Estos aportes metodológicos no tienen precedentes en la investigación psicométrica argentina.

Método

Participantes

La muestra estuvo conformada por dos grandes grupos: 1) estudiantes de la carrera de Psicología de la Facultad de Psicología de la Universidad de Buenos Aires (UBA) y 2) estudiantes de la Facultad de Arquitectura, Diseño y Urbanis-

mo de la UBA y estudiantes del primer año del Instituto Universitario Nacional del Arte (IUNA). El tamaño depurado de la primera muestra fue de 475 individuos, de los cuales aproximadamente 80% fueron mujeres y 20% varones. Estas proporciones son las usualmente encontradas en la población de estudiantes de Psicología de la UBA. La media y la mediana de edad fueron 21.87 y 20. En cuanto a la segunda muestra, su tamaño depurado fue de 1129 individuos, conformados en un 68% por estudiantes del Taller de Dibujo del primer año general de la UBA conocido como Ciclo Básico Común (CBC), en un 5% por estudiantes del IUNA, y el resto se repartía en los diversos niveles de la carrera de Arquitectura. En esta segunda muestra hubo un porcentaje mejor nivelado de hombres y mujeres (47% y 53%), una media y mediana de edad de 21.32 y 20. Las razones principales que llevaron a realizar el estudio con la segunda muestra fueron: 1) afinidad de los alumnos a los conceptos visuales y espaciales del TAF dado el contenido de las carreras, 2) nivelación de los porcentajes según el género y 3) superar las limitaciones del primer estudio realizado con la muestra de Psicología, a saber, a) el problema de la velocidad interviniendo en las respuestas, ya que en el primer estudio la correlación entre el tiempo total y el puntaje total resultó significativa al 1%, y b) el tamaño de la muestra acorde para un Modelo Logístico de Tres Parámetros (ML3P), ya que se recomienda que $n > 1000$ (Hanson & Beguin, 2002; Yen, 1987).

Procedimiento

Para la selección de los individuos, se estableció un diseño muestral no probabilístico, por accesibilidad o conveniencia (Gil Escudero y Martínez Arias, 2001). Los estudiantes fueron evaluados en las aulas del instituto donde se encontraban cursando, bajo su consentimiento y anónimos de su anonimato. El tiempo promedio que los individuos destinaron a la resolución del TAF fue de 42 minutos en ambas muestras.

Instrumentos

A cada uno de estos dos grandes grupos se le administró una versión diferente del TAF mediando la revisión de 15 ítems entre muestra y muestra; tal revisión está especificada en Blum, Galibert et al. (2011). Ambas versiones contuvieron una consigna inicial con tres ejercicios de práctica, más un protocolo con 36 reactivos figurales. Los ítems del protocolo de la versión revisada pueden consultarse en la página 42 de Blum, Galibert et al. (2011).

Para la construcción de los reactivos del protocolo, se utilizaron las reglas de rotación, traslación, distorsión del tamaño, distorsión de la forma, adición y sustracción. Se construyeron ítems con una regla, con dos reglas y con tres reglas, bajo la hipótesis de que el aumento de la cantidad de reglas elevaría la dificultad de los reactivos. Entre las pautas

utilizadas para su construcción (Blum, Abal, Lozzia et al., 2011), se cuidó especialmente que la respuesta correcta no pudiera encontrarse relacionando únicamente algunos elementos de la matriz. La razón de esto es que el motivo para elegir una respuesta debe basarse en el análisis previo de la estructura relacional de las figuras, es decir, $P(X,Y)$, y en la subsiguiente comparación entre pares relacionados $Q(X) \rightarrow Q(Y)$.

Resultados

Todos los análisis que se describirán a continuación se realizaron sobre tres matrices de datos. Las primeras dos matrices (M1 y M2) corresponden a las dos muestras respectivamente, mientras que la tercera matriz (M3) reúne los datos de los 1604 individuos de ambas muestras con respecto a los 21 ítems no modificados a través de las mismas.

Se analizó la unidimensionalidad de la prueba por medio de MicroFact 1.1 (Waller, 1995) y del uso conjunto de Tetcorr 2.1 (Enzmann, 2005) y SPSS 15, sobre las matrices de correlaciones tetracóricas. El supuesto de unidimensionalidad es muy importante, ya que los modelos logísticos unidimensionales de la TRI asumen su existencia. Además, a partir de la existencia de unidimensionalidad puede deducirse la presencia de independencia local, que es otro supuesto de la TRI. La unidimensionalidad se determinó a partir de los criterios siguientes: un porcentaje de varianza total explicada por el primer autovalor igual o mayor al 40% (Carmines & Zeller, 1979), la presencia del criterio de caída de Cattell (citado en García-Cueto y Fidalgo, 2005), una razón del primer autovalor al segundo $\lambda_1 / \lambda_2 > 5$ (Martínez Arias, 1995) y el ajuste global al ML3P de la TRI utilizando BILOG-MG (Zimowski, Muraki, Mislevy, & Bock, 1996), ya que logrando dicho ajuste con un modelo que supone la unidimensionalidad, también se comprueba la misma. Como M1 poseyó una distribución de frecuencias del puntaje total tendiente a la simetría (valor de asimetría: -0.10), se corrió BILOG asumiendo la distribución normal por defecto. En cambio, M2 y M3 poseyeron una distribución más asimétrica (valores de asimetría: -0.34 y -0.37 respectivamente). En estos dos casos y en función de la búsqueda de un mejor ajuste global, se corrió BILOG considerando una distribución a priori empírica.

En forma previa al análisis de componentes principales, se calculó el coeficiente Kaiser-Meyer-Olkin (*KMO*) y el Test de Esfericidad de Bartlett, obteniéndose resultados que indicaron la pertinencia de continuar con este análisis. Los resultados en cuanto a la unidimensionalidad fueron favorables considerando las tres matrices de datos, excepto por el ajuste en M3, tal como muestran la Tabla 1 y la Figura 1. Sin embargo, en este último caso debe recordarse que una muestra de 1604 individuos es muy grande y, como consecuencia, el estadístico de calidad de ajuste pasó a ser bastante más sensible al rechazo de H_0 .

Tabla 1. Análisis de la Unidimensionalidad.

	M1	M2	M3
Var. Explicada	39.37%	46.16%	44.85%
λ_1 / λ_2	8.38	12.22	9.53
Ajuste al ML3P	$\chi^2(276) = 276.8;$ $p = .47$	$\chi^2(293) = 308.6;$ $p = .25$	$\chi^2(188) = 279.4;$ $p < .001$

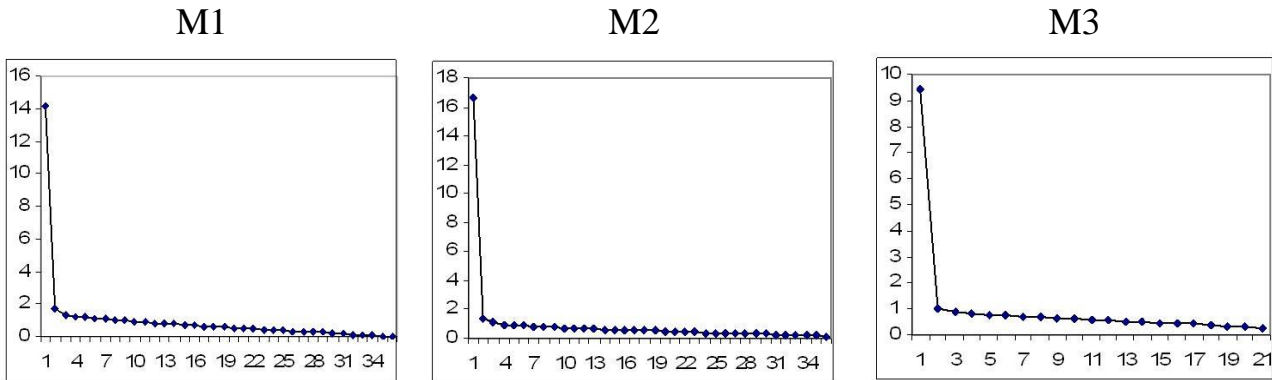


Figura 1. Diagramas de Autovalores.

Para el estudio desde la TCT se obtuvieron el Coeficiente a de Cronbach, Índices de Dificultad Corregidos (ID') e Índices de Discriminación basados en la correlación ítem-total corregida (r). Se utilizó también el programa BILOG-MG para determinar el ajuste ítem por ítem al ML3P de la TRI y estimar los parámetros de Discriminación (a), de Dificultad (b) y de Pseudoazar (c) de cada reactivo. Asimismo, se calculó la Función de Información del Test (FIT) y el Error Estándar de Estimación para aportar un criterio adicional a la confiabilidad de la prueba en función de los diferentes niveles de habilidad.

Se determinaron resultados muy favorables en cuanto a la confiabilidad, tanto desde la TCT al estudiar el a de Cronbach, como desde la TRI al estudiar las curvas que re-

presentan las Funciones de Información del Test (FIT). En particular, los estudios desde la TRI dieron cuenta de que la máxima información ($I. \text{Máx}$) quedó determinada en los puntos medios de la escala de habilidad, y que las FIT se aproximaron bastante a curvas esperables, es decir, simétricas respecto de $\theta = 0$ y una oscilación de la habilidad entre -3 y $+3$, tal como muestran la Tabla 2 y la Figura 2.

Tabla 2. Estudios de la Confiabilidad.

	M1	M2	M3
a	.91	.93	.88
$\theta I. \text{Máx}$	0.53	0.06	0.08

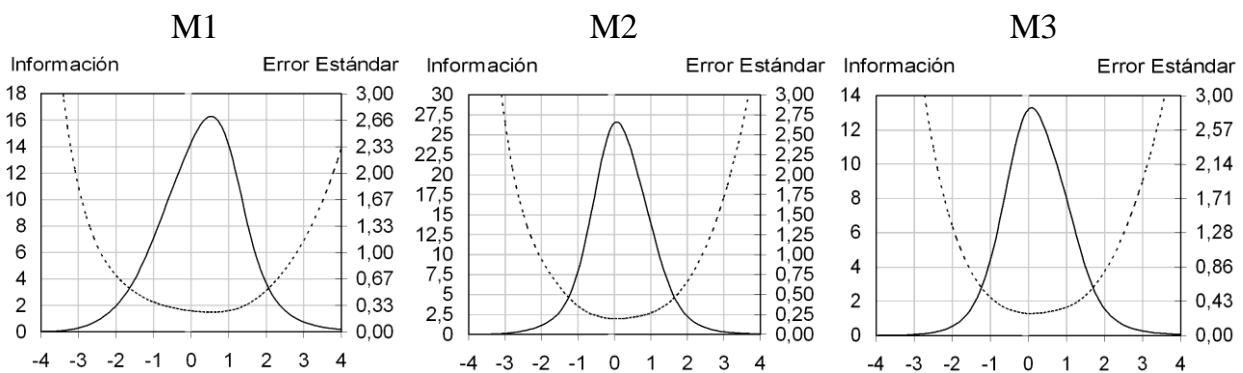


Figura 2. Funciones de Información del Test.

Según la información brindada en cuanto a los ID' y los r de la TCT, así como la estimación de los parámetros a , b y c de la TRI, se obtuvieron los siguientes resultados y se graficaron las siguientes CCI (Tabla 3 y Figura 3). A grandes rasgos, se obtuvieron índices y parámetros aceptables. En particular, todos los niveles de r puntuaron por encima de .30 y los niveles de a también fueron elevados sobre todo en la

segunda muestra. Los niveles de c poseyeron una puntuación esperable cuando se trabaja con 6 opciones de respuesta, es decir, alrededor de .17 o inferior. Como puede apreciarse, la dificultad media fue cercana a $\theta = 0$ y además, las CCI se agruparon mayormente hacia el centro representado por dicho nivel de habilidad. Esto quiere decir que la dificultad no fue muy variada. De hecho, los niveles máximos y mínimos

de b fueron: -1.77 y 1.13, -1.79 y 0.94, y -0.86 y 0.98 respectivamente para M1, M2 y M3. Existen otras investigaciones que prueban que pueden construirse ítems de analogía figural con niveles de b más variados. En una investigación similar (Embretson & Reise, 2000), los b fluctuaron entre -2.81 y 3.46 utilizando el ML3P. En la primera muestra, el promedio de los b de ítems con una regla difirió al 5% del promedio de b de ítems con dos reglas (-0.48 vs. 0.11. $t_{(28)} = -2.35$; $p = .03$. Tamaño del Efecto $TE = 0.93$), pero esta diferencia no fue significativa cuando se compararon ítems con dos reglas con los que poseen tres reglas (0.11 vs. 0.16. $t_{(25)} = -0.20$; $p = .84$). En vista de esta falta de diferencia y antes de realizar el segundo estudio, los ítems que poseían tres reglas se modificaron para que contuvieran solo dos de ellas. Luego de admi-

nistrarlos a la segunda muestra, se obtuvo que la comparación de los b de ítems con una y con dos reglas brindó datos significativos al 5% (-0.42 vs. 0.01. $t_{(34)} = -2.37$; $p = .02$. $TE = 0.91$).

Tabla 3. Índices de la TCT y Parámetros de la TRI.

	M1	M2	M3
\overline{ID}^p	.49	.54	.53
\bar{r}	.45	.50	.48
\bar{a}	1.02	1.26	1.17
\bar{b}	-0.03	-0.09	-0.06
\bar{c}	.14	.15	.15

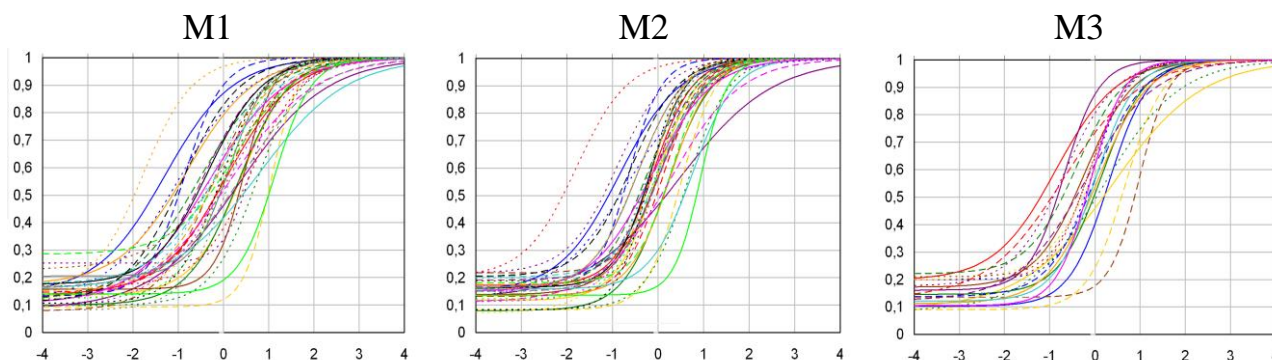


Figura 3. Curvas Características de los Ítems.

Se establecieron regresiones múltiples en cada muestra, tomando como variable criterio al parámetro b y, como potenciales predictores, las seis reglas utilizadas en este estudio. Se escogió aquel modelo de regresión cuyo C_p de Mallows (1973) fuera el más bajo. Como medidas del ajuste y del tamaño del efecto de la regresión múltiple se utilizaron respectivamente el Coeficiente de Determinación (r^2) y F^2 de Cohen, cuya fórmula es $r^2 / (1 - r^2)$. Se encontró lo siguiente al 5%. En M1, el mejor modelo de regresión ($C_p = 2.8$) identificó a la rotación y a la sustracción como predictores de b ($F_{(2,33)} = 5.43$; $p < .01$), mientras que en M2, el mejor modelo de regresión ($C_p = 5.9$) determinó a la rotación, la traslación, la distorsión del tamaño, la adición y la sustracción como

predictores de b ($F_{(5,30)} = 2.94$; $p = .03$). La varianza conjunta explicada y los F^2 de Cohen son, en M1, 24.77% y 0.33 y en M2, 32.90% y 0.49.

A continuación se muestran otros resultados de estas regresiones (ver Tabla 4). Puede apreciarse que todos los coeficientes parciales son positivos, con lo cual, la presencia de las reglas mencionadas predice el aumento de b . Por otro lado, las reglas de emplazamiento espacial contribuyeron más al aumento de la dificultad que las reglas de distorsión. Este último hallazgo es consistente con los resultados de Whitely y Schneider (1981) según los cuales es esperable que los ítems con cambios de emplazamiento espacial resulten más difíciles que los ítems con distorsiones.

Tabla 4. Regresiones Múltiples.

	Predictores	Coefficientes de Regresión	Coefficientes de Determinación Parcial	Contribución a la Varianza
Mejor modelo en M1	Rotación	.62	.44	19.26%
	Sustracción	.53	.40	16%
Mejor modelo en M2	Rotación	.52	.42	17.96%
	Traslación	.52	.42	17.51%
	Dist. Tamaño	.39	.35	12.31%
	Adición	.31	.30	9.21%
	Sustracción	.73	.54	29.02%

Finalmente, en la Tabla 5 se muestran todos los índices y parámetros estudiados y se incluye el estudio del ajuste de cada ítem al ML3P. Cada fila representa los datos de un

mismo ítem, sólo que se modificó el orden de presentación de algunos reactivos en M2 y por eso su numeración es distinta respecto de M1. Como puede observarse, la gran mayo-

ría de los reactivos ajustó al ML3P en las tres matrices de datos. Los ítems que no ajustaron en M1 son el 5, el 11 y el 18, en M2 son el 6, el 11, el 15 y el 17, y en M3 son el 16, el 23 y

el 27. Por lo tanto, sólo el ítem 11 no ajustó en más de una matriz de datos, lo que lo hace más proclive a la eliminación (Blum, Auné, Galibert y Attorresi, 2013).

Tabla 5. Índices, Parámetros y Estadísticos de Bondad de Ajuste χ^2 .

M1							M2							M3						
It.	ID'	r	b	a	c	χ^2 (p)	It.	ID'	r	b	a	c	χ^2 (p)	ID'	r	b	a	c	χ^2 (p)	
1	.47	.45	0.08	0.81	.14	8.5 (.49)	1	.57	.53	-0.10	1.35	.17	3.4 (.91)							
2	.78	.39	-1.29	0.79	.14	5.8 (.44)	2	.72	.45	-0.85	0.88	.15	8.0 (.43)	.74	.43	-0.86	0.87	.197	11.6 (.24)	
3	.37	.48	0.40	1.08	.14	3.0 (.96)	3	.53	.56	-0.04	1.32	.12	6.4 (.61)							
4	.36	.52	0.33	1.05	.10	8.9 (.45)	4	.41	.55	0.20	1.23	.08	7.9 (.45)	.40	.53	0.30	1.35	.10	5.7 (.77)	
5	.37	.39	0.34	0.61	.08	24.5 (.004)	5	.39	.32	0.47	0.57	.12	7.2 (.62)	.38	.34	0.39	0.59	.09	13.3 (.15)	
6	.35	.56	0.46	1.81	.16	7.4 (.28)	30	.51	.51	0.04	1.19	.15	7.1 (.63)							
7	.47	.46	0.06	0.88	.14	3.2 (.95)	7	.48	.54	0.10	1.36	.13	14.5 (.10)	.48	.50	0.14	1.23	.15	12.0 (.21)	
8	.75	.53	-0.86	1.40	.13	2.2 (.70)	8	.74	.58	-0.63	1.67	.17	12.1 (.10)	.74	.56	-0.68	1.51	.16	12.8 (.12)	
9	.18	.49	1.05	1.95	.09	6.5 (.60)	9	.30	.54	0.50	1.55	.08	8.1 (.42)							
10	.60	.44	-0.53	0.75	.12	11.5 (.18)	10	.60	.52	-0.23	1.18	.16	5.8 (.67)	.60	.50	-0.23	1.09	.17	14.9 (.09)	
11	.61	.48	-0.44	0.91	.13	16.8 (.01)	11	.56	.51	-0.05	1.26	.19	15.7 (.047)							
12	.53	.53	-0.25	0.98	.08	7.5 (.49)	36	.59	.55	-0.21	1.34	.15	8.9 (.35)							
13	.34	.44	0.62	1.03	.17	6.8 (.66)	19	.42	.51	0.27	1.33	.14	10.7 (.30)							
14	.70	.41	-0.98	0.75	.12	8.5 (.29)	14	.63	.41	-0.61	0.71	.12	10.7 (.30)	.65	.40	-0.70	0.69	.13	15.9 (.07)	
15	.45	.50	0.08	0.93	.12	7.4 (.50)	15	.54	.53	-0.21	1.03	.07	17.4 (.03)	.51	.50	0.01	1.12	.14	11.8 (.22)	
16	.24	.49	0.73	1.16	.10	5.8 (.76)	16	.26	.48	0.64	1.31	.08	5.3 (.73)	.25	.47	0.67	1.41	.09	20.2 (.02)	
17	.37	.44	0.67	1.78	.25	7.4 (.49)	17	.78	.47	-0.89	1.08	.22	15.8 (.04)							
18	.54	.35	0.06	0.74	.23	20.5 (.01)	24	.66	.53	-0.44	1.18	.16	5.1 (.75)							
19	.69	.40	-0.81	0.79	.17	4.2 (.76)	31	.65	.52	-0.46	1.15	.16	11.7 (.17)	.67	.48	-0.68	1.02	.18	9.3 (.41)	
20	.34	.32	0.78	0.60	.16	2.6 (.98)	20	.29	.40	0.75	1.24	.16	16.2 (.06)							
21	.54	.42	-0.31	0.71	.10	8.7 (.47)	21	.55	.45	-0.13	0.89	.15	7.5 (.59)	.55	.43	-0.41	1.09	.22	4.0 (.91)	
22	.17	.37	1.13	1.44	.14	4.5 (.87)	22	.18	.38	0.94	1.99	.14	7.8 (.55)	.18	.34	-0.20	0.78	.12	16.4 (.06)	
23	.61	.48	-0.33	0.97	.18	1.4 (.98)	23	.59	.57	-0.16	1.50	.17	1.1 (.997)	.60	.53	0.98	1.79	.14	29.8 (.001)	
24	.51	.51	0.14	1.39	.21	3.7 (.72)	18	.55	.59	0.01	1.88	.19	5.6 (.70)							
25	.53	.47	-0.12	0.94	.15	7.0 (.54)	25	.59	.58	-0.13	1.66	.18	5.1 (.75)	.57	.53	-0.11	1.44	.198	9.6 (.38)	
26	.54	.43	-0.10	0.84	.18	6.0 (.65)	26	.54	.48	0.05	1.20	.203	1.6 (.99)	.54	.44	-0.03	1.53	.20	6.5 (.69)	
27	.43	.39	0.15	0.64	.11	13.3 (.15)	27	.41	.41	0.26	0.76	.11	4.8 (.85)	.42	.40	0.04	1.01	.199	21.9 (.01)	
28	.53	.36	0.25	0.87	.28	7.8 (.46)	28	.45	.51	0.30	1.49	.18	5.4 (.79)							
29	.71	.47	-0.90	0.93	.11	10.2 (.11)	29	.71	.49	-0.58	1.13	.204	2.9 (.94)	.71	.48	0.19	0.70	.09	10.3 (.33)	
30	.40	.37	0.45	0.71	.16	5.9 (.75)	12	.54	.49	0.10	1.35	.22	8.7 (.37)							
31	.90	.37	-1.77	1.07	.15	2.7 (.44)	13	.90	.39	-1.79	1.04	.203	11.3 (.08)							
32	.56	.48	-0.21	1.01	.16	7.3 (.40)	32	.60	.54	-0.14	1.49	.198	7.9 (.44)	.59	.51	-0.09	1.37	.21	12.6 (.18)	
33	.44	.46	0.06	0.77	.09	16.5 (.06)	33	.50	.51	-0.02	1.09	.11	4.9 (.77)	.48	.50	0.05	1.07	.11	12.1 (.21)	
34	.47	.55	0.03	1.19	.12	2.8 (.91)	34	.53	.55	-0.07	1.31	.12	7.2 (.52)	.51	.54	0	1.32	.12	14.3 (.11)	
35	.49	.57	-0.06	1.21	.11	2.8 (.90)	35	.57	.65	-0.21	1.70	.08	3.6 (.74)	.54	.60	-0.11	1.59	.11	14.3 (.11)	
36	.42	.59	0.11	1.19	.08	7.3 (.51)	6	.51	.51	0.01	1.14	.14	25.4 (.001)							

Discusión

Dificultades encontradas y su posible solución

Más allá de los avances alcanzados en la elaboración de la prueba, aún quedan por desarrollar perfeccionamientos. De aquí en más, debe controlarse mejor la complejidad de las figuras y es necesario emplear criterios específicos para confeccionar las opciones de respuesta. Esto se debe a que, entre otras cuestiones, existieron problemas con el diseño de ítems con tres reglas en M1. Se esperaba que tales reactivos fueran más difíciles de resolver que los ítems con dos reglas; sin embargo, la diferencia entre los *b* de ambos grupos de ítems no fue significativa al 5%. Inspeccionando los ítems con tres reglas, se confirmó que muchos de ellos podían contestarse en forma correcta con sólo reconocer dos de

esas reglas, lo cual es una consecuencia de la forma de plantear sus opciones de respuesta. Tales opciones deberían ser suficientemente parecidas como para que el evaluado encuentre la opción correcta con ayuda del empleo de todas las reglas que participan en el ítem, y no de algunas. Eliminar una regla de cada uno de los reactivos con tres reglas permitió simplificar los ítems en cuestión y reducir su complejidad frente al segundo estudio, pero esto implicó dejar de trabajar con tres reglas para concentrar la atención sólo en ítems más sencillos de construir. En algunos ítems podrían separarse mejor las opciones más parecidas y siguen existiendo cuestiones por modificar basándose en el análisis de los distractores.

Todo esto permite pensar en un fuerte componente humano influyendo negativamente en la construcción de ítems. Si bien la mayor parte de las medidas que se mencionarán a

continuación se han tomado en cuenta, la rigurosidad de su definición y estructuración es necesaria para mejorar sustancialmente la sistematización del proceso de construcción.

- 1- Reducir la complejidad al mínimo necesario es una norma ineludible. La complejidad añade información al reactivo, información que debe ser procesada por el respondiente y que lo somete a una mayor probabilidad de cometer error (Mulholland, Pellegrino, & Glaser, 1980). Como el propósito no es medir esta complejidad sino la variable de interés (el RA o la Inteligencia General), debe diseñarse el reactivo sólo con los pocos elementos que servirán para aplicar las reglas. De ser posible y para ahorrar elementos, deberían aplicarse varias reglas a las mismas partes de la figura (Freund et al., 2008).
- 2- Debe existir mayor sutileza de las diferencias entre las figuras. Martínez, Moreno y Muñoz (2005) determinaron que las opciones de respuesta deberían ser parecidas entre sí. Para los fines aquí propuestos, esto se traduce en buscar la sutileza del cambio entre las figuras de la matriz, lo que permitirá elaborar opciones ligeramente distintas unas de otras. Por ejemplo, si se emplea una regla de rotación, tal vez en ciertos casos convenga rotar una parte de la figura y no toda; si se aplica la sustracción como regla, debe sustraerse un fragmento pequeño. De este modo, el cambio será más difícil de observar y el acierto del ítem dependerá del grado de concentración invertido y del nivel de habilidad de la persona.
- 3- Con el ideal de construir reactivos con una, dos y/o más reglas que regulen adecuadamente la dificultad, se deberían elaborar opciones de respuesta por medio de un Diseño de Combinación de Alternativas (DCA). Entiéndase por 'alternativa correcta' a una solución basada en la aplicación adecuada de una regla en el ítem, y por 'alternativa incorrecta' a una solución basada en su aplicación inadecuada. El constructor debe desarrollar, para cada regla que se presenta en el ítem, una sola alternativa correcta y una o más alternativas incorrectas. Luego le es posible combinar una alternativa de cada regla para generar una opción de respuesta, pudiendo realizar la totalidad de combinaciones disponibles para crear todas las opciones del ítem. La única respuesta correcta será aquella opción que contenga todas las alternativas correctas. Para un ejemplo práctico, analícese en detalle los tres ítems mostrados en la Figura 4. El número de opciones, por defecto, es igual al producto de los números que representan la cantidad propuesta de alternativas por regla; sin embargo, podría trabajarse con menos opciones al eliminar los distractores más alejados de la respuesta correcta, que son aquellas opciones que poseen la mayor cantidad de alternativas incorrectas. Al realizar este proceso, el constructor debe ser cuidadoso y no eliminar todas las alternativas incorrectas que pertenecen a alguna de las reglas, es decir, el ítem debe conservar al menos una alternativa incorrecta por regla a lo largo de las opciones no eliminadas. De lo contrario, se estarían elimi-

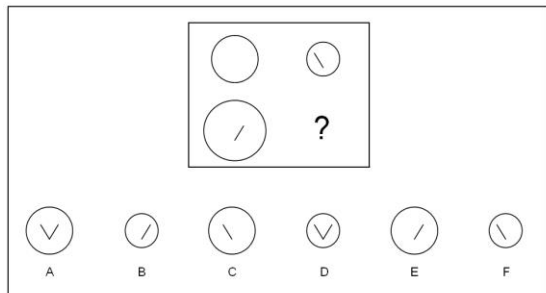
nando los puntos de comparación que permiten resolver el ítem sobre la base de esa o esas reglas.

Pautas específicas para la construcción de ítems

Es posible desglosar las tres sugerencias mencionadas y conformar sugerencias específicas, planteadas a modo de pautas o pasos para la construcción manual de ítems con matrices de 2x2. Estos pasos se detallarán a continuación, utilizando un lenguaje similar al pseudocódigo para la creación de un programa de computadora, es decir, como si la intención real fuera manipular artificialmente los radicales (Irvine, 2002) de un Ítem-Modelo (Gierl & Lai, 2012; Lai, Alves, & Gierl, 2009) en un proceso de Generación Automática de Ítems (Arendasy, 2002; Freund, Hofer, & Holling, 2008). Los pasos presuponen que un ítem debe contener entre 2 y 8 opciones de respuesta, ya que cantidades mayores a 8 crean ítems con una estructura demasiado compleja y es posible que dicha complejidad distorsione las respuestas.

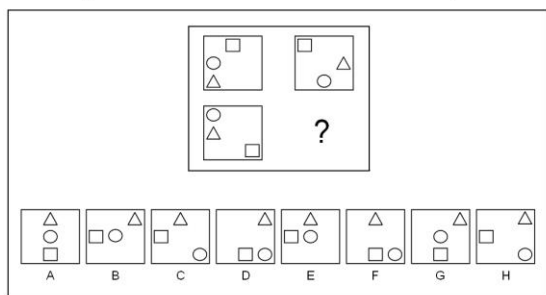
- 1- Proponer una cantidad de reglas: pensar en un valor del 1 al 3.
- 2- Asignar a cada regla una operación cognitiva elemental (tipo de regla): ejemplos de operaciones cognitivas encontradas en la literatura son rotación, traslación, reflejo, distorsión del tamaño, distorsión de la forma, adición y sustracción. Pueden proponerse otras, aunque éstas tal vez son más que suficientes. Una descripción de las reglas mencionadas puede encontrarse en Blum, Abal, Lozzia et al. (2011).
- 3- Elaborar una figura inicial: se confecciona la figura de partida, la cual se emplazará en el costado superior izquierdo de la matriz. Es aconsejable que la figura no contenga demasiada complejidad, para evitar que la complejidad contamine la respuesta al ítem.
- 4- Descomponer la figura: debe indicarse de qué manera hay que descomponer la figura para que pueda asignarse una regla por cada una de las partes en que la misma se descompuso. De este modo, cada regla va a actuar independientemente en cada parte designada de la figura. Si se trabaja con una sola regla, igualmente debe indicarse qué parte de la figura se verá afectada por dicha regla.
- 5- Designar la cantidad de alternativas por regla, considerando lo siguiente:
 - Si se quiere trabajar con una sola regla, la cantidad mínima de alternativas es 2 y pueden utilizarse hasta 8 alternativas.
 - Si se quiere trabajar con dos reglas, pueden utilizarse (a) dos alternativas por regla, (b) dos alternativas para una regla y tres para la otra regla, o (c) tres alternativas por regla.
 - Si se quiere trabajar con tres reglas, la única cantidad de alternativas admisible por regla es 2.

Ítem Tamaño-Adición. Alternativas: 2x3. Opciones: 6.



		Alternativas	
Reglas		Correcta	Incorrectas
Tamaño			
Adición			

Ítem Triple Traslación. Alternativas: 2x2x2. Opciones: 8.



		Alternativas	
Reglas		Correcta	Incorrecta
Traslación			
Traslación			
Traslación			

Ítem Triple Rotación. Alternativas: 2x2x2. Opciones: 8.

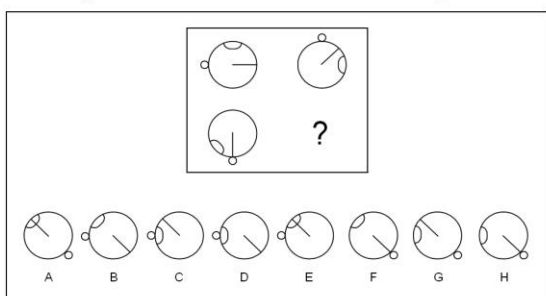


Figura 4. Diseño de Combinación de Alternativas en Ítems de Matrices Figurales de 2x2.

- 6- Designar la cantidad de opciones de respuesta. El número de opciones está restringido a la cantidad de reglas y de alternativas por regla. Tomar en cuenta lo siguiente:
- Si la cantidad de reglas es 1, el número de opciones es indefectiblemente igual al número de alternativas.
 - Si la cantidad de reglas es 2 y se quiere trabajar con dos alternativas por regla, la cantidad de opciones por defecto es 4, pero el investigador puede elegir sólo 3 opciones.
 - Si la cantidad de reglas es 2 y se quiere trabajar con dos alternativas para la primera regla y tres para la segunda, la cantidad de opciones por defecto es 6, pero pueden designarse 5.

- Si la cantidad de reglas es 2 y se quiere trabajar con tres alternativas por regla, la cantidad de opciones máxima sugerida es 8, pero pueden escogerse 7.
 - Si la cantidad de reglas es 3, la cantidad de opciones por defecto es 8, pero puede reducirse a 7, 6, 5 o 4.
- 7- Designar el número o código de orden, de izquierda a derecha, en que se emplazará la respuesta correcta, o establecer aleatoriamente dicho orden.

Todos los pasos mencionados asientan las bases a priori para construir una matriz de 2x2 y un conjunto de opciones de respuesta debajo de aquélla, tal como se ve en los reacti-

vos de la Figura 4. En términos específicos, el investigador debe procesar dichos pasos en dos etapas:

- 1- Construcción de la matriz: requiere de la previa lectura de los pasos 1, 2, 3 y 4 explicadas arriba. El investigador toma la figura inicial, la descompone según los criterios especificados y aplica una regla por cada parte previamente identificada para realizar alteraciones. Se vale de este procedimiento para generar dos figuras modificadas respecto de la figura inicial. La forma de aplicar estas reglas la define el investigador. Una de las figuras modificadas es colocada hacia el costado derecho respecto de la figura inicial, y la otra es colocada debajo de la misma. Siempre quedará vacío un cuarto espacio (ver 'signo de pregunta'). De este modo se genera la matriz de 2x2 como se ve en los ejemplos.
- 2- Construcción de las opciones de respuesta: requiere de la previa lectura de todos los pasos ya explicados. El investigador desarrolla, por cada regla, una sola alternativa correcta y versiones de alternativas incorrectas. Luego combina una alternativa de cada regla para generar una opción de respuesta posible o, en el caso de trabajar con una sola regla, utiliza una alternativa para generar una opción de respuesta posible. Así, genera todas las opciones de respuesta realizando la cantidad total disponible de combinaciones. Luego emplaza las opciones uniformemente, en hilera horizontal y orden aleatorio, y ubica la respuesta correcta (aquella opción cuyas alternativas son todas correctas) en el orden designado. Considera lo siguiente:
 - Si se trabaja con la cantidad de opciones sugerida por defecto, el investigador combina las alternativas para formar las opciones sin realizar un proceso de eliminación de opciones. La excepción la tiene el caso donde se proponen dos reglas y tres alternativas por regla, ya que si bien hay nueve opciones que resultan de la combinación, el investigador va a generar un máximo de ocho opciones al eliminar al azar una opción cuyas alternativas son todas incorrectas.
 - Si se trabaja con dos reglas y con una cantidad de opciones menor a la sugerida por defecto, el investigador primero combina las alternativas para formar las opciones y luego elimina una o la única opción cuyas alternativas son todas incorrectas. Si proponiendo dos reglas y tres alternativas por regla, la cantidad de opciones a designar es 7, entonces del total de nueve combinaciones se eliminan dos opciones cuyas alternativas deben ser todas incorrectas.
 - Si se trabaja con tres reglas y con una cantidad de opciones menor a la sugerida por defecto, existe un orden de prioridad para la eliminación. De acuerdo a la cantidad de opciones propuesta por el investigador, primero se elimina aquella opción cuyas alternativas son todas incorrectas y luego, si corresponde, se eliminan al azar una, dos o todas las opciones que contengan una sola alternativa correcta y otras dos alternativas incorrectas. Las demás opciones nunca se eliminan.

Una vez procesado el ítem, el archivo final debe contener:

- 1- Información escrita:
 - El número o código de orden de la respuesta correcta.
 - Nombres de cada regla utilizada.
- 2- Un cuadro rectangular que contenga al ítem en cuestión, tal como se ve en los ejemplos. Es preferible guardarlo en formato de dibujo de alta calidad, por ejemplo un Metarchivo Mejorado (Enhanced Metafile, EMF) de Windows.

También podría considerarse la posibilidad de realizar todo este proceso con cuatro reglas, aunque el diseño se volvería demasiado complejo en este caso. La única cantidad admisible de alternativas por regla sería 2, la cantidad mínima admisible de opciones sería 5 (una respuesta correcta y cuatro distractores, cada uno de los cuales contendría una sola alternativa incorrecta y tres correctas) y la cantidad máxima de opciones sería 8 ya que existiría la posibilidad de añadir distractores con dos alternativas correctas y dos incorrectas cada uno.

Conclusiones

Para la medición del Razonamiento Analógico u otros componentes de la Inteligencia General, muchas veces existe el inconveniente de que se utilizan pruebas muy convencionales, conocidas popularmente, que se publican en sitios web y están al alcance de todas las personas. Si bien la construcción de nuevas escalas se presenta como algo viable, es sabido que lleva mucho tiempo y esfuerzo elaborar una prueba en forma manual y que a su vez la misma presente las propiedades psicométricas esperadas. La lectura de material conceptual y la experiencia pueden servir de apoyo para dicha construcción, pero siempre se presentan dificultades basadas en el factor humano, las cuales obligan a realizar correcciones.

Es importante mencionar que existe una demanda creciente de miles de nuevos ítems para incluir en evaluaciones computarizadas por parte de agencias de evaluación muy conocidas (Gierl & Lai, 2012). Por ende, a medida que el desarrollo de ítems se complejiza, la demanda de cantidad y calidad de ítems excede con creces la capacidad de aquellos redactores que los arman del modo usual (Lai, Alves, & Gierl, 2009). Por lo tanto, la proposición de pautas específicas para la construcción de ítems, con escasos desarrollos en Argentina, permitiría crear reactivos de la manera más sistemática y objetiva posible, sin cometer grandes errores. Con el mismo fin, se están desarrollando actualmente estudios dentro del campo de la Generación Automática de Ítems, pero estas investigaciones son todavía muy recientes y su contribución a la construcción de tests con buenas propiedades psicométricas aún está puesta en duda. En un futuro cercano, tal vez los progresos en dicha temática permitan acortar gradualmente la brecha que existe entre la investigación psicométrica y los avances de la tecnología.

Agradecimientos.- Se agradece la participación de los Lic. en Psicología: Romina Bershady, Susana Cappellini, Lidia Nicolai y Ezequiel Katz en la recolección de datos de la segunda muestra.

Referencias

- Abad, F., Colom, R., Rebollo, I., & Escorial, S. (2004). Sex differential item functioning in the Raven's Advanced Progressive Matrices: evidence for bias. *Personality and Individual Differences*, *36*, pp. 1459-1470.
- Arendasy, M. (2002). *GeomGen—Ein Itemgenerator für Matrixtestaufgaben*. Wien: Eigenverlag.
- Arendasy, M. (2005). Automatic generation of Rasch-calibrated items: figural matrices test GEOM and Endless-Loops Test EC. *International journal of testing*, *5*(3), 197-224.
- Bandeira Andriola, W. (2000). Funcionamento Diferencial dos Itens (DIF): Estudo com Analogias para Medir o Raciocínio Verbal. *Psicologia: Reflexão e Crítica*, *13*(3), pp. 475-483.
- Blum, G. D., Abal, F. J. P., Galibert, M. S. y Attorresi, H. F. (2011). Construcción de una Prueba de Analogías Figurales. *Summa Psicológica UST*, *18*(1), 5-12.
- Blum, G. D., Abal, F. J. P., Lozzia, G.S., Picón Janciro, J. C. y Attorresi, H. F. (2011). Analogías de figuras: Teoría y construcción de ítems. *Interdisciplinaria. Revista de psicología y ciencias afines*, *28*(1), 131-144.
- Blum, G. D., Auné, S., Galibert, M. S. y Attorresi, H. F. (2013). Criterios para la eliminación de ítems de un Test de Analogías Figurales. *Summa Psicológica UST*, *10*(2), 49-56.
- Blum, G. D., Galibert, M. S., Abal, F. J. P., Lozzia, G. S. y Attorresi, H. F. (2011). Modelización de una Prueba de Analogías Figurales con la Teoría de Respuesta al Ítem. *Escritos de Psicología*, *4*(3), 36-43.
- Brown, L., Sherbenou, R. J. y Johnsen, S. K. (2000). *TONI 2. Test de Inteligencia No Verbal. Apreciación de la habilidad cognitiva sin influencia del lenguaje. Manual*. Madrid: TEA.
- Carmine, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment*. Londres: Sage.
- Cattell, R. B., & Cattell, A. K. S. (1973). *Measuring Intelligence with The Culture Fair Tests*. Champaign, IL: Institute for Personality and Ability Testing, Inc.
- Cubillo, J. C. y González Labra, M. J. (1998). El razonamiento analógico como solución de problemas. En M. J. González Labra (Ed.), *Introducción a la psicología del pensamiento* (pp. 409-451). Madrid: Trotta.
- De la Fuente Aranz, J. y Minervino, R. A. (2004). Pensamiento analógico. En M. Carretero & M. Asensio (Coords.), *Psicología del pensamiento* (pp. 193-214). Madrid: Alianza.
- Embretson, S. E., & Reise, S. P. (2000). *Item Response Theory for psychologists*. Mahwah: Lawrence Erlbaum Associates Inc.
- Enzmann, D. (2005). *Dirk Enzmann - Statistical Software (Some Useful Things)*. Recuperado de http://www2.jura.uni-burg.de/instkrim/kriminologie/Mitarbeiter/Enzmann/Software/Enzmann_Software.html.
- Freund, P. A., Hofer, S., & Holling, H. (2008). Explaining and controlling for the psychometric properties of computer-generated figural matrix items. *Applied Psychological Measurement*, *32*(3), pp. 195-210.
- García-Cueto, E. y Fidalgo, A. M. (2005). Análisis de los ítems. En J. Muñiz, A.M. Fidalgo, E. García-Cueto, R. Martínez, & R. Moreno (Eds.), *Análisis de ítems* (pp. 53-130). Madrid: La Muralla.
- Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, *7*(2), pp. 155-170.
- Gierl, M.J., & Lai, H. (2012). The role of item models in automatic item generation. *International journal of testing*, *12*(3), 273-298. Doi: 10.1080/15305058.2011.635830.
- Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive psychology*, *12*(3), pp. 306-355.
- Gil Escudero, G. y Martínez Arias, M. R. (2001). Metodología de encuestas. En M. J. Navas (Ed.), *Métodos, diseños y técnicas de investigación psicológica* (pp. 379-436). Madrid: Universidad Nacional de Educación a Distancia.
- Hanson, B. A., & Beguin, A. A. (2002). Obtaining a common scale for item response theory item parameters using separate versus concurrent estimation in the common-item equating design. *Applied Psychological Measurement*, *26*(1), pp. 3-24.
- Holyoak, K. J., & Koh, K. (1987). Surface and structural similarity in analogical transfer. *Memory and cognition*, *15*(4), pp. 332-340.
- Holyoak, K. J., & Thagard, P. R. (1989). A computational model of analogical problem solving. En S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 242-266). Cambridge University Press.
- Irvine, S. (2002). The foundations of item generation for mass testing. En S.H. Irvine, & P.C. Kyllonen (Eds.), *Item generation for test development* (pp. 3-34). Mahwah: Lawrence Erlbaum Associates.
- Kubinger, K. D. (2008). On the revival of the Rasch model-based LLTM: from constructing tests using item generating rules to measuring item administration effects. *Psychology science quarterly*, *50*(3), 311-327.
- Lai, H., Alves, C., & Gierl, M. J. (2009). Using automatic item generation to address item demands for CAT. En D.J. Weiss (Ed.), *Proceedings of the 2009 GMAC Conference on Computerized Adaptive Testing*. Recuperado de: www.psych.umn.edu/psylabs/CATCentral.
- Maller, S. J. (2000). Item invariance in four subtests of the Universal Nonverbal Intelligence Test (UNIT) across groups of deaf and hearing children. *Journal of psychoeducational assessment*, *18*(3), pp. 240-254.
- Mallows, C. L. (1973). Some Comments on *Cp. Technometrics*, *15*(4), 661-675. Doi:10.2307/1267380.
- Martínez Arias, R. (1995). *Psicometría: Teoría de los Tests Psicológicos y Educativos*. Madrid: Síntesis.
- Martínez, R., Moreno, R. y Muñiz, J. (2005). Construcción de los ítems. En J. Muñiz, A.M. Fidalgo, E. García-Cueto, R. Martínez, & R. Moreno (Eds.), *Análisis de los ítems* (pp. 9-48). Madrid: La Muralla.
- Mulholland, T. M., Pellegrino, J. W., & Glaser, G. (1980). Components of geometric analogy solution. *Cognitive Psychology*, *12*(2), pp. 252-284.
- Raven, J., Raven, J. C., & Court, J. H. (1991). *Manual for Raven's Progressive Matrices and Vocabulary Scales. Sections 1, 2, 3 and 4*. Oxford: Oxford Psychologists Press.
- Sternberg, R. J. (1977). *Intelligence, information processing and analogical reasoning: the componential analysis of human abilities*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Sternberg, R. J. (1987). *Inteligencia humana II: Cognición, personalidad e inteligencia*. Barcelona: Paidós.
- Waller, N. G. (1995). *MicroFact 1.1. A Microcomputer Factor Analysis Program for Ordered Polytomous Data and Mainframe Size Problems*. St. Paul Minnesota: Assessment System Corporation.
- Wechsler, D. (1997). *Wechsler Adult Intelligence Scale-III*. Nueva York: Psychological Corporation.
- Whitely, S. E., & Schneider, L. M. (1981). Information structure for geometric analogies: A test theory approach. *Applied Psychological Measurement*, *5*(3), pp. 383-397.
- Wolf Nelson, N., & Gillespie, L. L. (1991). *Analogies for thinking and talking. Words, pictures and figures*. Tucson: Communication Skill Builders.
- Yen, W. M. (1987). A comparison of the efficiency and accuracy of BILOG and LOGIST. *Psychometrika*, *52*(2), pp. 275-291.
- Zeuch, N. (2010). *Rule-based item construction: analysis with and comparison of linear logistic test models and cognitive diagnostic models with two item types* (Tesis doctoral, Westfälische Wilhelms-Universität Münster, Germany). Recuperado de http://miami.uni-muenster.de/servlets/DerivateServlet/Derivate-6086/diss_zeuch.pdf.
- Zimowski, M., Muraki, E., Mislevy, R., & Bock, R. (1996). *BILOG-MGTM: Multiple-group IRT analysis and test maintenance for binary items* [Software]. Chicago, IL: Scientific Software International.

(Artículo recibido: 20-5-2013; revisado: 16-01-2013; aceptado: 10-11-2013)