

Reinhardt Fischer,
Gerardo Prieto¹
Universidad de Salamanca, España

Aunque el EAT-26 es un test muy utilizado mundialmente en el diagnóstico e investigación de los trastornos alimenticios, apenas existen estudios de sus propiedades psicométricas, utilizando el Modelo de Rasch (RM). El objetivo principal de este estudio ha sido analizar, con el Modelo de Escalas de Calificación, un modelo tipo Rasch para ítems politómicos, el test EAT-26 en una muestra de adolescentes. Así mismo, se propuso una adaptación de este test, con el fin de utilizarlo en el contexto paraguayo. La reagrupación de las 6 categorías originales y la eliminación de ítems que no se ajustaban al modelo derivaron en la propuesta de una versión del EAT, con 18 ítems y 3 categorías de respuesta. Se aportan los resultados correspondientes a las características de la nueva versión: unidimensionalidad, ajuste al modelo de personas e ítems, confiabilidad de las medidas, diferencias en la escala entre hombres y mujeres, y funcionamiento diferencial de los ítems asociado al sexo.

Palabras clave: Trastornos alimenticios, Modelo de Rasch, Modelo de Escalas de Calificación, Test de Actitudes Alimentarias.

Although the EAT-26 is a test widely used worldwide in the diagnosis and research of eating disorders, there are hardly any studies of its psychometric properties using Rasch Model (RM). The first objective of this study was to analyze with the Rating Scale Model, an extension of RMI for polytomous items, EAT-26 test on a sample of adolescents. The second objective was to propose an adaptation of the EAT-26 for its use in the Paraguayan context. Collapsing the 6 original categories and eliminating items that do not fit the model, a version of the EAT -with 18 items and 3 response categories- was proposed. The results for the characteristics of the new version are provided: unidimensionality, fit to the model of people and items, reliability, differences in scale between men and women and differential item functioning related with sex.

Keywords: Eating disorders, Rasch Model, Rating Scale Model, Eating Attitudes Test.

Recibido: 26 de julio de 2016

Aprobado: 30 de Agosto de 2016

1. La correspondencia puede ser dirigida a Reinhardt Fischer, Universidad de Salamanca, Departamento de Psicología Básica, Psicobiología y Metodología, Avda. de la Merced 109-131, 37005, Salamanca (España). E-mail: lemat_fischer@hotmail.com

Los trastornos del comportamiento alimentario (TCA) son considerados como enfermedades psiquiátricas graves, marcadas por alteraciones en el comportamiento, las actitudes y la ingestión de alimentos, y generalmente acompañadas de intensa preocupación por el peso o por la forma del cuerpo (Portela de Santana, da Costa Ribeiro Junior, Mora Giral, & Raich, 2012)".

Los TCA constituidos por la anorexia nerviosa (AN), la bulimia nerviosa (BN) y el TCA no especificado, representan en la actualidad uno de los problemas de salud más importante, tanto por el número creciente de personas afectadas, como por las muertes que ocasiona (Baldares, 2013).

Muchos estudios sobre la incidencia de los TCA se centran en presentar datos originarios de los servicios de salud, por lo que resulta difícil establecer la prevalencia/incidencia en la población general (Portela de Santana et al., 2012). Por ello, el uso de instrumentos ágiles que identifiquen las personas con riesgo de TCA contribuye al diagnóstico precoz y al desarrollo de programas de prevención. Cabe anotar que uno de los instrumentos más usados es el Test de Actitudes Alimentarias (Eating Attitudes Test) (EAT) (Constaín et al., 2014).

El test fue diseñado por Garner y Garfinkel en 1979, con el propósito de desarrollar y validar una escala de calificación que puede ser útil en la evaluación de una amplia gama de comportamientos y actitudes que se encuentran en la anorexia nerviosa (Garner & Garfinkel, 1979). En un comienzo, el cuestionario contaba con 40 ítems (EAT-40). Sin embargo, los estudios acerca de sus cualidades psicométricas condujeron a la abreviación del mismo (Gouveia, de Lucena Pronk, Santos, Gouveia, & Cavalcanti, 2010), resultando así un cuestionario de 26 ítems (EAT-26).

Existen varias versiones del EAT-26. Por ejemplo, la versión EAT-18 de seis factores propuesta por Maiano, Morin, Lanfranchi, & Therme (2013), que quizás sea el trabajo de adaptación más ambicioso hecho hasta el momento. Lo anterior debido a que incluyó una serie de cuatro estudios con una muestra grande ($n = 2178$) y heterogénea, en la que participaron adolescentes provenientes de distintos contextos (europeos, africanos, americanos y asiáticos). Cabe anotar que Gouveia et al. (2010) proponen una versión abreviada (EAT-12), compuesta por tres factores, para un contexto español; otros estudios, como el de Orbitello et al. (2006), hacen referencia a la validez del EAT-26, pero con un punto de corte distinto al del original.

La mayoría de estos estudios de adaptación se centran en análisis derivados de la Teoría Clásica de los Tests (TCT). Aunque la TCT se ha mostrado útil para muchos propósitos, presenta algunas limitaciones tales como la dependencia: las propiedades psicométricas de los ítems dependen de la muestra de personas en la que han sido obtenidas y, asimismo, las puntuaciones de las personas dependen del conjunto de ítems integrado en el instrumento de diagnóstico (Muñiz, 2010). Además, la métrica de los tests clásicos no suele superar el nivel ordinal y, en el caso de las escalas de tipo Likert, el RM permite analizar empíricamente la funcionalidad de las categorías de respuesta (Prieto & Delgado, 2007).

Este tipo de limitaciones en el análisis pueden ser subsanadas con modelos de medida invariante como el modelo de Rasch (1960) y sus extensiones. Estos modelos poseen propiedades como la objetividad específica (los parámetros de las personas no dependen de la muestra de ítems y los parámetros de los ítems son independientes de la muestra de personas) y la medición conjunta: si los datos se ajustan al modelo, se puede medir a las personas y a los ítems en una misma dimensión con propiedades de intervalo (Prieto & Delgado, 2003).

Para analizar ítems politómicos, como los del test EAT-26, se puede emplear uno de los modelos tipo Rasch: el Modelo de Escalas de Calificación (MEC), que fue propuesto por Andrich (1978). Una ventaja del MEC es que permite analizar empíricamente la calidad métrica de las opciones de respuesta.

En consecuencia, este trabajo ha tenido como principal objetivo analizar con el MEC el test EAT-26 en una muestra de adolescentes, con la finalidad de obtener un instrumento apropiado para su uso diagnóstico en el contexto paraguayo.

En esta investigación, se utilizó un enfoque metodológico cuantitativo, de tipo instrumental, el cual está encaminado a desarrollar pruebas, incluyendo tanto el diseño (o adaptación) como el estudio de las propiedades psicométricas de los mismos (Montero & León, 2007).

En este estudio se utilizó una muestra no probabilística, compuesta por 210 participantes (100 de sexo masculino y 110 de sexo femenino), con edades comprendidas entre 14 y 19 años. Todos ellos eran estudiantes de bachillerato, seleccionados de tres colegios de la ciudad de Bella Vista Sur, Paraguay.

El Test de Actitudes Alimentarias (Eating Attitudes Test, o EAT-26) (Garner, Olmsted, Borhr, & Garfinkel, 1982). Para responder a este instrumento, el participante debe leer los ítems e indicar con qué frecuencia ha realizado cada uno de los comportamientos expresados en los mismos. Para ello, debe utilizar una escala tipo Likert de seis puntos: 1 = Nunca, 2 = Rara vez, 3 = A veces, 4 = Frecuentemente, 5 = Casi siempre, 6 = Siempre.

Una psicóloga clínica se encargó de aplicar los cuestionarios en los distintos colegios. El ingreso a las instituciones educativas se gestionó a través de solicitudes dirigidas a los directivos de las mismas. La aplicación se realizó en las aulas de clase, en horario cedido por los profesores. Aunque aplicados en ambiente colectivo, los instrumentos fueron respondidos individualmente. La psicóloga presentó las mismas instrucciones que se indican en el manual del test a todos los sujetos, poniendo énfasis en el carácter voluntario de la participación y la garantía de que toda la información era confidencial.

El análisis de los datos se realizó con el programa WINS-TEPS (Linacre & Wright, 2011), utilizando el Modelo de Escalas de Calificación, una extensión del modelo de Rasch para ítems politómicos. Este modelo puede ser representado con la siguiente fórmula:

$$\ln (P_{nik} / P_{ni(k-1)}) = B_n - D_i - F_k$$

P_{nik} = probabilidad de que el sujeto n responda al ítem i con la categoría k

$P_{ni(k-1)}$ = probabilidad de que el sujeto n responda al ítem i con la categoría $k-1$

B_n = parámetro en el rasgo del sujeto n

D_i = parámetro en el rasgo del ítem i

F_k = paso entre las categorías k y $k-1$

Se asume que los pasos entre las categorías (F_k) son iguales en todos los ítems.

Los valores de los parámetros se expresan en una unidad de medida especial denominada *logit*. Aunque la escala logit puede adoptar valores entre más y menos infinito, en la gran mayoría de los casos se sitúa en el rango ± 5 . La localización del punto 0 de la escala es arbitraria, situándose en dicho punto la dificultad media de los ítems (Prieto & Velasco, 2003).

La precisión de los estimadores de los parámetros se cuantifica mediante el error estándar, que es la desviación típica de los errores al estimar el parámetro de la persona o del ítem. Además, es posible cuantificar la confiabilidad global en la muestra de personas o de ítems mediante los estadísticos denominados *confiabilidad de las medidas de los sujetos* (PSR, por el nombre en inglés Person Separation Reliability) y *confiabilidad de las medidas de los ítems* (ISR, por el nombre en inglés Item Separation Reliability). La PSR y la ISR, que pueden oscilar entre 0 y 1, tienen una interpretación análoga al coeficiente de fiabilidad clásico: la proporción de la varianza observada que no está asociada a la varianza de error.

Para analizar el ajuste de personas e ítems al modelo, se tuvo en cuenta la media no ponderada de los residuos estandarizados al cuadrado (Outfit) y la media de los residuos estandarizados al cuadrado ponderados con su valor de información (Infit). Cuando los valores de estos indicadores oscilan entre 0.5 y 1.5 se considera que los datos tienen un ajuste adecuado al modelo; entre 1.5 y 2.0, el desajuste es considerado moderado y sin graves consecuencias para las medidas, mientras que los valores mayores a 2.0 indican un desajuste severo (Linacre, 2011).

Adicionalmente al empleo de un análisis ómnibus del ajuste al modelo (Infit y Outfit), se ha generalizado en los últimos años la utilización de contrastes específicos del supuesto básico del modelo: la unidimensionalidad. Este supuesto suele examinarse mediante el Análisis de Componentes Principales (ACP) de los residuos (Chou & Wang, 2010). Dada la falta de consenso sobre los criterios que sean indicativos de una dimensión secundaria (Chou & Wang, 2010), se consideraron como criterios de una unidimensionalidad esencial los siguientes: que el porcentaje de la varianza explicado por las medidas Rasch fuese al menos superior al 20% (Reckase, 1978), que el autovalor del primer componente de los residuos fuese inferior a 3.0 y que el porcentaje de varianza explicada por ese primer componente fuese inferior al 10% (Chou & Wang, 2010; Miguel et al., 2013).

Para evaluar la calidad de las categorías se utilizaron las reglas de Linacre (2002):

1. Al menos 10 observaciones en cada categoría: para estimar los pasos (puntos de transición) entre las categorías, es necesario una cantidad mínima de elecciones en cada categoría.
2. Distribución regular de las observaciones (aceptables las distribuciones unimodales o bimodales). Inadecuadas las distribuciones en dientes de sierra: la distribución óptima es la que tiene el mismo porcentaje de observaciones en cada categoría, pero se admiten; por ejemplo, muchas elecciones en una categoría extrema y que las siguientes vayan en descenso.
3. Incremento monótonico de los promedios de las categorías: según van ascendiendo las categorías, tiene que ir ascendiendo el nivel en el atributo medido.
4. Outfit menor de 2.0: el ajuste de las categorías se estima de igual forma que el ajuste de personas e ítems.
5. Incremento monótonico de los pasos entre las categorías: por ejemplo, el paso entre las categorías 2 y 3 debe ser mayor que el paso entre las categorías anteriores, 1 y 2, y así, sucesivamente.

Finalmente, dado que su presencia representa una amenaza a la validez de las puntuaciones, se llevó a cabo la detección del Funcionamiento Diferencial de los Ítems (DIF) asociado al sexo. La comparación de los parámetros de dificultad de los hombres y de las mujeres se utilizó como técnica de detección del DIF (Prieto & Nieto, 2014). Wright y Douglas (1976) observaron que las diferencias inferiores a .50 logit no tuvieron ninguna consecuencia para la

validez de las medidas. Por lo tanto, se detecta la presencia de DIF si la diferencia entre los parámetros es superior a .50 logit. La prueba t con ajuste de Bonferroni (Benjamini y Hochberg, 1995) se utilizó para probar la significación y se describe a continuación.

$$t = Df - Dr / \sqrt{(SEDf^2 + SEDr^2)/2}$$

Es decir, SEDf y SEDr son los errores estándar de los dos parámetros de dificultad. Si alguna de las pruebas t de la lista tiene $p < .05 / (\text{número de pruebas t en la lista})$, entonces se rechaza la hipótesis de no DIF (corrección de Bonferroni).

El primer análisis se llevó a cabo con las 6 categorías originales del test (1 = nunca, 2 = rara vez, 3 = a veces, 4 = frecuentemente, 5 = casi siempre, 6 = siempre).

Como se observa en la Tabla 1, no hay un incremento monótonico de los pasos entre las categorías, lo que resulta en el aplanamiento de las curvas de algunas de ellas. Si una curva de probabilidad es aplanada, la categoría no es funcional puesto que no es la de más probable elección en ningún rango de la variable (Figura 1).

Se procedió a reagrupar las categorías centrales de la siguiente manera: la categoría 2 se conformó de la unión de las categorías 2 y 3; y la categoría 3 la conformó la unión de las categorías 4 y 5. Entonces, esto dio como resultado cuatro opciones de respuesta (1 = nunca, 2 = a veces, 3 = con frecuencia, 4 = siempre).

En la Figura 1 se observa que, con esta nueva agrupación, aún no hay un incremento monótonico de los pasos.

Tabla 1
Estadísticos de las seis categorías originales

N= 210	1	2359 (43)	-1.11	.95	.99	----
----	2	986 (18)	-.61	1.16	1.07	-.00
----	3	955 (17)	-.37	.98	.87	-.51
----	4	360 (7)	-.16	1.01	1.02	.69
----	5	316 (6)	.02	1.04	1.01	.07
----	6	484 (9)	.22	1.11	1.27	-.25

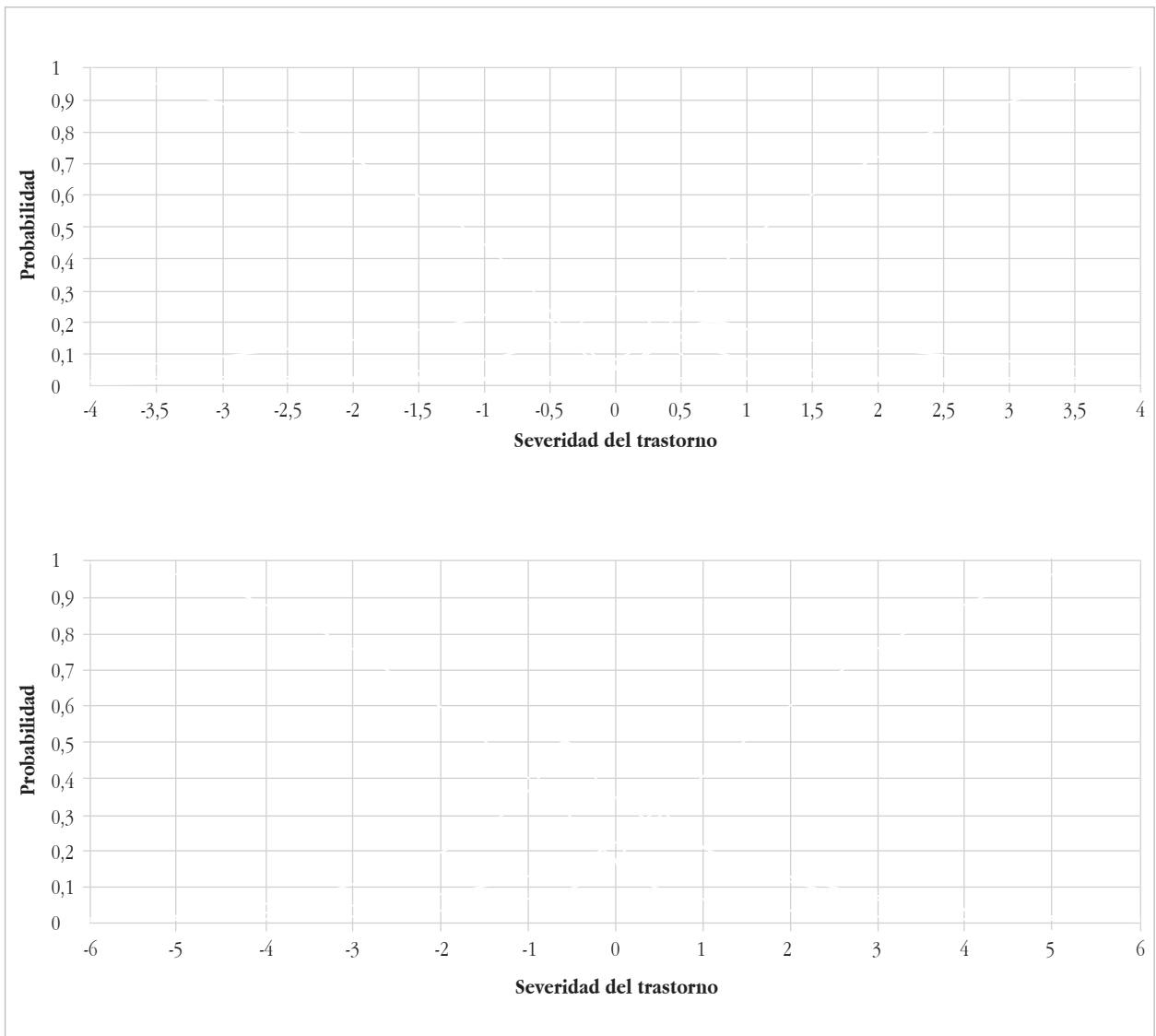


Figura 1.
Curvas características de los ítems con seis y cuatro categorías

Así bien, se llevó a cabo una nueva agrupación: la categoría 2 se conformó de la unión de las antiguas categorías 2 y 3, lo que dio como resultado tres categorías de respuesta

(1= casi nunca, 2= en ocasiones, 3= casi siempre). Con esta reagrupación, como se presenta en la Tabla 2, se cumplen todas las reglas de Linacre (2002).

Tabla 2
Estadísticos de las tres categorías finales

N= 210	1	3345 (61)	-1.87	1.04	1.17	----
----	2	1315 (24)	-.64	.97	.81	-.30
----	3	800 (15)	.25	1.01	1.02	.30

Para investigar el ajuste de la prueba al supuesto de unidimensionalidad, se realizó un Análisis de Componentes Principales (ACP) de los residuos.

En el EAT-26, el porcentaje de la varianza explicado por la medida Rasch fue del 38.4%, valor que supera netamente el 20% mínimo indicado por Reckase (1979) como un primer requisito de unidimensionalidad. Sin embargo, el primer componente de los residuos tenía un valor propio de 3.1, por lo que no cumple con la recomendación de que no exceda un valor de 3.0 (Chow & Wang, 2010; Miguel, Silva & Prieto, 2013). Con la finalidad de obtener una medida fundamentalmente unidimensional, se procedió a eliminar los ítems con menor homogeneidad (una correlación con la variable inferior a .30). De esta forma, se eliminaron 7 ítems (Tabla 3).

Una vez excluidos los 7 ítems, se llevó a cabo un nuevo ACP de los residuos. El porcentaje de la varianza explicado por la medida Rasch fue del 39%; así, el primer componente tenía un valor propio de 2.5, lo que corresponde a un 7.9% de la varianza explicada, cumpliendo de esta forma con los criterios de unidimensionalidad establecidos.

Tabla 3
Ítems eliminados por baja homogeneidad

4	He tenido atracones de comida en los cuales siento que no puedo parar de comer	.21
8	Siento que los demás quieren que yo coma más	.25
9	Vomito después de haber comido	.26
13	Los demás piensan que estoy muy delgado(a)	.21
21	Pienso demasiado en la comida	.28
25	Disfruto probando nuevas comidas apetitosas	.16
26	Tengo el impulso de vomitar luego de comer	.18

Además del análisis de la dimensionalidad de la escala, se examinó la existencia de funcionamiento diferencial del ítem (DIF) asociado al sexo. Se encontró que el ítem 2 (Evito comer cuando tengo hambre) presentaba DIF, dado que los parámetros de localización en la muestra de hombres (.09) y en el de las mujeres (1.06) diferían en casi 1 logit. La diferencia era significativa ($p = .0008$). Los datos manifiestan que la conducta que se describe en el ítem indica mayor severidad del trastorno en las mujeres que en los hombres. Por consiguiente, se optó por eliminar ese ítem en la versión final de la escala EAT que, por incluir 18 ítems, denominaremos EAT-18. En esta versión, el porcentaje de la varianza explicado por la medida Rasch fue del 39,1% y el primer componente tenía un valor propio de 2.4, lo que corresponde a un 8.3% de la varianza explicada.

En la Tabla 4 se observan los estadísticos de los ítems de la versión final EAT-18, el ajuste (Infit y Outfit), la localización (D) y el error estándar (SE).

Los valores de Infit de los ítems varían entre .74 y 1.39, mientras que los valores de Outfit en 17 de los 18 ítems oscilan entre .68 y 1.49. Por tanto, los valores de ajuste de casi todos los ítems se incluyen en el rango óptimo para medir adecuadamente. Sólo un ítem presenta un Outfit superior a 1.50 (1.68), lo que indica un desajuste moderado que no daña la validez de las medidas (Linacre, 2011).

Los parámetros de localización de los ítems varían entre 1.40 y -1.38, un rango moderadamente alto que garantiza una evaluación de la severidad aceptable. A los ítems 6 y 18 (ambos con 1.40 logit) y al ítem 3 (-1.38 logit) les corresponde los extremos en severidad (en el Apéndice puede consultarse el contenido de los ítems). La distribución de los ítems a lo largo de la variable TCA puede apreciarse en la Figura 2.

Dado que la muestra de personas tiene un tamaño adecuado, se ha estimado con una buena precisión la dificultad de los ítems: los errores estándar son bajos y el coeficiente de confiabilidad ($ISR = .98$), elevado.

Se ha de notar, asimismo, que los índices clásicos de discriminación (correlación ítem-test) varían entre .31 y .65, por lo que se considera que la prueba tiene una homogeneidad adecuada.

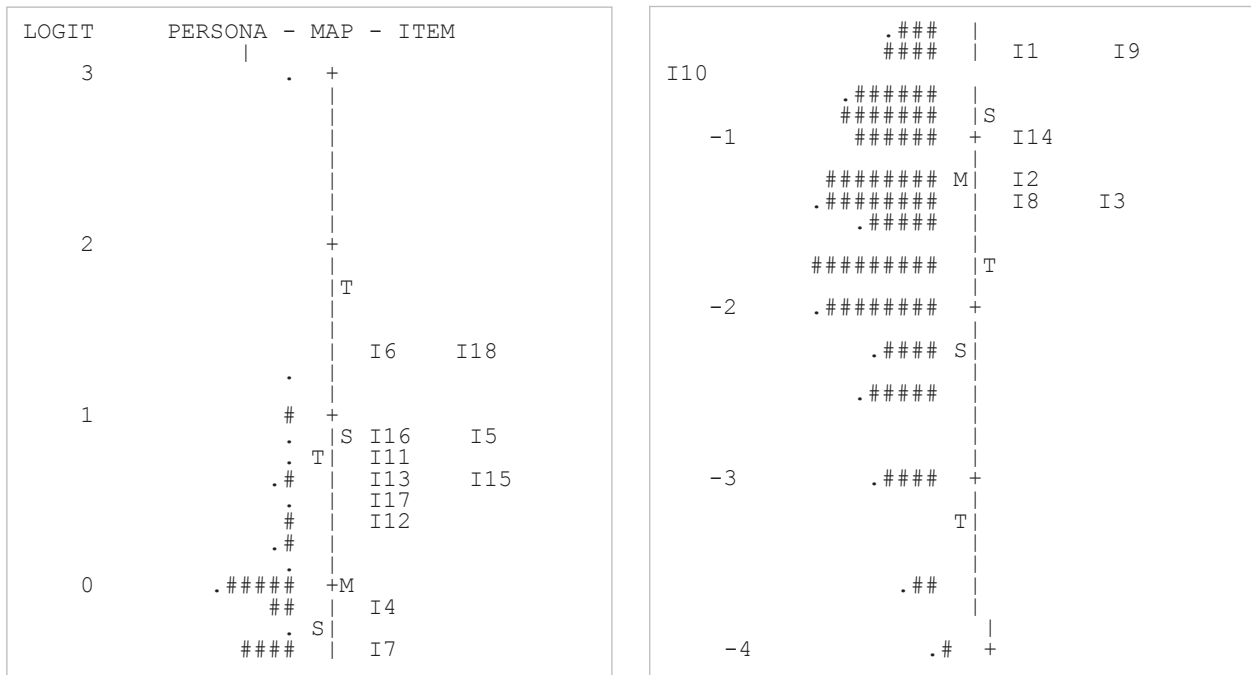


Figura 2. Distribución de personas (izquierda) e ítems (derecha) a lo largo del continuo TCA. Personas e ítems por encima del punto cero muestran mayor severidad en TCA

Tabla 4

Estadísticos de los ítems: ajuste (*Infit* y *Outfit*); la localización (*d*); el error estándar (*SE*) y homogeneidad

1	1.01	.98	-.56	.11	.53
2	1.04	1.10	-1.22	.10	.45
3	1.35	1.49	-1.38	.10	.37
4	.84	.74	-.16	.12	.58
5	.87	.71	.92	.15	.47
6	1.22	.82	1.40	.18	.41
7	.97	1.03	-.37	.11	.51
8	.84	.80	-1.37	.10	.63
9	.74	.73	-.58	.11	.65
10	1.17	1.25	-.63	.11	.44
11	1.09	1.13	.74	.15	.38
12	.82	.70	.33	.13	.52
13	1.26	1.23	.58	.14	.40
14	1.08	1.07	-1.04	.10	.47
15	1.39	1.63	.60	.14	.31
16	.89	.71	.85	.15	.52
17	.82	.68	.50	.14	.54
18	.99	.79	1.40	.18	.42
Media	1.02	.98	.00	.13	--
SD	.19	.28	.90	.03	--

En la Tabla 5 se incluyen los estadísticos descriptivos de los valores de las personas.

Cabe mencionar que se destacan los estadísticos de ajuste. Las medias de Infit y Outfit apenas se separan de 1 (ajuste perfecto). La variabilidad de los valores de ajuste es baja: sólo el 2.4% de las personas presentan un desajuste severo con el modelo (Outfit > 2). Se trata de un porcentaje muy bajo que permite concluir que el ajuste de las personas al modelo es aceptable.

En lo respecta al nivel de severidad de los trastornos alimenticios de las personas evaluadas, se observa que la media de las puntuaciones logit (-1.22) es netamente inferior a la media de los ítems (el punto cero de la escala), lo que sugiere que la mayor parte de la muestra tiene un nivel de severidad bajo en la variable medida. No obstante, la variabilidad de las personas es elevada (DT = 1.01).

Se ha de destacar que la fiabilidad de las puntuaciones logit de las personas en EAT-18 es adecuada (PSR = .76). De acuerdo con Evers et al. (2013), se considera que la fiabilidad es adecuada cuando PSR oscila entre .70 y .80; buena, entre .80 y .90; y excelente si PSR es mayor de .90.

Tabla 5
Resumen de los estadísticos descriptivos de las personas

Infit de personas	2.46	.34	1.01	.36
Outfit de personas	2.97	.29	.98	.48
Localización de las personas	2.95	-3.68	-1.22	1.01
Localización varones	1.20	-3.68	-1.39	.97
Localización mujeres	2.95	-3.68	-1.06	1.02

En la Tabla 6 y la Figura 2 se muestra la distribución de las personas en la variable medida. La distribución de las puntuaciones de las personas es asimétricamente positiva (un número de personas moderadamente bajo presenta altos niveles de severidad). En la Tabla 6 también aparecen unas normas de grupo en percentiles.

Otro detalle distinguible en la Tabla 5 es que la puntuación máxima obtenida por las mujeres (2.95) supera ampliamente a la obtenida por los varones (1.20). En promedio, las mujeres

presentaron una mayor severidad (-1.06) que los hombres (-1.39). La diferencia es estadísticamente significativa (t de Welch = 2.03; $gl = 207$; $p = .04$) y el tamaño del efecto es medio-bajo (d de Cohen = -.33).

El propósito central de la presente investigación ha sido analizar el test EAT-26 con el Modelo de Escalas de Calificación en una muestra de adolescentes paraguayos.

En primera instancia, se encontró que las 6 categorías de respuesta originales del test no funcionan de manera adecuada, por lo que se llevaron a cabo varias agrupaciones de las categorías adyacentes, hasta lograr un sistema de 3 categorías métricamente apropiado de acuerdo con las recomendaciones de Linacre (2002).

Dado que las ventajas del modelo Rasch sólo pueden ser obtenidas si los datos empíricos se ajustan al modelo (Prieto & Delgado, 2003), se investigó mediante el Análisis de Componentes Principales (ACP) de los residuos, si el EAT-26 cumplía con el supuesto básico de unidimensionalidad. Puesto que el primer componente residual tenía un valor propio por encima del recomendado, se procedió a eliminar ítems con una baja correlación ítem-test, resultando así un EAT con 19 ítems. Cabe destacar que la varianza total explicada por la medida Rasch no se alteró de forma sustancial tras la eliminación de los 7 reactivos. Asimismo, se eliminó un ítem que presentaba funcionamiento diferencial asociado al sexo. El re-análisis mediante el ACP de la escala con los 18 ítems restantes mostró que la versión que proponemos presenta una unidimensionalidad esencial en el sentido de que la probabilidad de las respuestas de las personas a los ítems depende de la localización de las personas y los ítems en una dimensión básica. De hecho, los estadísticos de ajuste de los ítems y de las personas, basados en las medias de los residuos (diferencias entre las respuestas observadas y las predichas por el modelo), corroboran que es suficiente una dimensión para medir conjuntamente la severidad del trastorno de las personas y de las conductas. La conclusión referente a la unidimensionalidad no coincide con otros estudios en los que se describen distintas estructuras factoriales de diversas versiones del EAT (Garner et al., 1982; Gouveia et al., 2010; Maiano et al., 2015). La diversidad de resultados puede deberse al distinto número de ítems de las versiones, al método de análisis factorial utilizado (exploratorio o confir-

Tabla 6
Distribución de las puntuaciones de las personas y normas en percentiles

18	-4.89E	1.83	3	1
19	-3.68	1.01	5	3
20	-2.96	.72	9	6
21	-2.53	.60	11	11
22	-2.21	.53	9	15
23	-1.95	.49	17	22
24	-1.74	.45	18	30
25	-1.54	.43	11	37
26	-1.37	.41	17	44
27	-1.20	.40	16	51
28	-1.05	.39	12	58
29	-.90	.38	14	64
30	-.76	.37	13	71
31	-.63	.36	8	76
32	-.50	.36	7	79
33	-.37	.36	8	83
34	-.24	.35	1	85
35	-.12	.35	4	86
36	.01	.35	11	90
37	.13	.35	1	93
38	.26	.35	3	94
39	.38	.36	2	95
40	.51	.36	1	95
41	.64	.36	3	96
42	.77	.37	1	97
43	.91	.37	1	98
44	1.05	.38	2	99
45	1.20	.39	1	99
46	1.36	.41	0	99
47	1.54	.43	0	99
48	1.73	.45	0	99
49	1.95	.48	0	99
50	2.20	.53	0	99
51	2.52	.60	0	99
52	2.95	.72	1	99

matorio), a los distintos criterios de extracción y rotación de factores, además del método y la teoría usada para el análisis. Sin embargo, en varios de estos estudios, son muy elevadas las correlaciones entre las dimensiones extraídas, lo que indica que en ellas subyace una dimensión fundamental que legitima el uso de modelos TRI.

En suma, en función de los resultados obtenidos, se podría asumir que los ítems tienden a articularse en torno a una sola dimensión, y parecería sensato asumir que el EAT-18 es un instrumento adecuado para estimar mediante una única puntuación el grado de TCA de las personas.

En cuanto a la estimación de los parámetros de la versión final del test, los resultados indican que tanto los ítems, como las personas se estimaron con adecuada fiabilidad.

La distribución de las personas en la escala logit mostró que la mayoría de los sujetos presentan un bajo nivel en la variable trastornos alimenticios; dicha distribución también pudo ser estimada en una escala en percentiles. Cabe mencionar que las normas en percentiles son útiles para localizar el nivel de severidad de una persona en referencia a su grupo normativo.

El análisis de impacto asociado al sexo reveló que aunque el tamaño del efecto es medio-bajo, el nivel medio en severidad de los trastornos alimenticios es superior en las mujeres que en los hombres. Este dato concuerda con el grueso de teorías (Fernández & Erro, 2004; Portela de Santana et al., 2012), lo que adquiere relevancia dado que en este estudio se evaluó a una muestra con una cantidad similar de varones y mujeres. Por el contrario, en la mayor parte de los trabajos (Constaín et al., 2014; Gouveia et al., 2010; Nunes, Camey, Olinto, & Mari, 2005), los participantes eran mayoritariamente del sexo femenino.

El presente trabajo permite proponer una versión del EAT que resulta apropiada para el diagnóstico de los trastornos alimentarios en muestras de adolescentes paraguayos. En la evaluación de los TCA, son indispensables las pruebas de cribado. Cabe anotar que la finalidad de estas pruebas es facilitar la detección de manifestaciones provocadas por la alteración de alguna función superior mediante escalas de fácil y rápida aplicación (Mora-Simón et al., 2012). La disponibilidad de estos instrumentos en la clínica le permitiría al psicólogo tener una idea general de la situación del paciente. Además de la utilidad clínica, los cuestionarios de cribado se utilizan en las investigaciones o en los programas dirigidos a la detección temprana de una

enfermedad en la población general (Pineda, Guerrero, Pinilla, & Estupiñán, 2002). Por esta razón, es necesario analizar los instrumentos de rastreo tan ampliamente utilizados, como el EAT, con modelos psicométricos avanzados, como el modelo Rasch, con muestras representativas de los contextos en los que se pretenden utilizar. En este trabajo se propone que la versión reducida del test (el EAT-18 con 3 categorías de respuesta) tiene unas propiedades psicométricas adecuadas y podría ser empleada como test de cribado en una población de adolescentes y jóvenes en Paraguay. Con este fin, se ha incluido en el apéndice la versión definitiva de la escala. No obstante, se requiere desarrollar estudios adicionales con muestras de sujetos clínicos y controles sanos, con el fin de determinar puntos de corte que faciliten la clasificación diagnóstica.

Andrich, D. (1978). Application of a psychometric rating model to ordered categories which are scored with successive integers. *Applied psychological measurement*, 2(4), 581–594.

Baldares, M. J. V. (2013). Trastornos de la Conducta Alimentaria. *Revista Médica de Costa Rica y Centroamérica*, 70(607), 475–482.

Benjamini, Y., Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1), 289-300.

Chou, Y.-T., & Wang, W.-C. (2010). Checking Dimensionality in Item Response Models With Principal Component Analysis on Standardized Residuals. *Educational and Psychological Measurement*, 70(5), 717-731. <http://doi.org/10.1177/0013164410379322>

Constaín, G. A., Ricardo, C., Rodríguez-Gázquez, M. A., Álvarez M., Marín, C., & Agudelo, C. (2014). Validez y utilidad diagnóstica de la escala EAT-26 para la evaluación del riesgo de trastornos de la conducta alimentaria en población femenina de Medellín, Colombia. *Atención Primaria*, 46(6), 283-289. <http://doi.org/10.1016/j.aprim.2013.11.009>

Evers, A., Hagemester, C., Hostmaelingen, A., Lindley, P., Muñoz, J., & Sjöberg, A. (2013). EFPA Review Model for the description and evaluation of psychological and educational tests. *Test review form and notes for reviewers. Version 4.2. 6*. Brussels: European Federation of Psychology Associations. Recuperado a partir de <http://www.academia.edu/download/38683296/>

- Fernández, L., & Erro, G. (2004). *Trastornos de la alimentación: de la evaluación al tratamiento*, Madrid: UNED Ediciones.
- Garner, D. M., & Garfinkel, P. E. (1979). The Eating Attitudes Test: an index of the symptoms of anorexia nervosa. *Psychological Medicine*, *9*, 273-279.
- Garner, D. M., Olmsted, M. P., Borhr, Y., & Garfinkel, P. E. (1982). The Eating Attitudes Test: psychometric features and clinical correlates. *Psychological Medicine*, *11*, 871-878.
- Gouveia, V. V., de Lucena Pronk, S., Santos, W. S., Gouveia, R. S., & Cavalcanti, J. P. (2010). Test de Actitudes Alimentarias: Evidencias de Validez de una Nueva Versión Reducida. *Interamerican Journal of Psychology*, *44*(1), 28–36.
- Linacre, J. M. (2002). Optimizing rating scale category effectiveness. *Journal of Applied Measurement*, *3*(1), 85–106.
- Linacre, J. M., & Wright, B. D. (2011). *A user's guide to WINS-TEPS: Rasch-model computer program*. Chicago: Mesa Press.
- Maiano, C., Morin, A. J. S., Lanfranchi, M.-C., & Therme, P. (2015). Body-related Sport and Exercise Motives and Disturbed Eating Attitudes and Behaviours in Adolescents: Body-Related Sport and Exercise Motives. *European Eating Disorders Review*, *23*(4), 277-286. <http://doi.org/10.1002/erv.2361>
- Miguel, J. P., Silva, J. T., & Prieto, G. (2013). Career Decision Self-Efficacy Scale — Short Form: A Rasch analysis of the Portuguese version. *Journal of Vocational Behavior*, *82*(2), 116-123. <http://doi.org/10.1016/j.jvb.2012.12.001>
- Montero, I., León, O. G. (2007). A guide for naming research studies in Psychology. *International Journal of Clinical and Health Psychology*, *7*(3), 847-862.
- Mora-Simón, S., García-García, R., Perea-Bartolomé, M. V., Ladera-Fernández, V., Unzueta-Arce, J., Patino-Alonso, M. C., & Rodríguez-Sánchez, E. (2012). Deterioro cognitivo leve: detección temprana y nuevas perspectivas. *Revista de Neurología*, *54*(5), 303–310.
- Muñiz, J. (2010). Las teorías de los tests: teoría clásica y teoría de respuesta a los ítems. *Papeles del psicólogo*, *31*(1), 57–66.
- Nunes, M. A., Camey, S., Olinto, M. T. A., & Mari, J. de J. (2005). The validity and 4-year test-retest reliability of the Brazilian version of the Eating Attitudes Test-26. *Brazilian Journal of Medical and Biological Research*, *38*(11), 1655–1662.
- Orbitello, B., Ciano, R., Corsaro, M., Rocco, P. L., Taboga, C., Tonutti, L., Balestrieri, M. (2006). The EAT-26 as screening instrument for clinical nutrition unit attenders. *International journal of obesity*, *30*(6), 977–981.
- Pineda, D. A., Guerrero, O. L., Pinilla, M. L., & Estupiñán, M. (2002). Utilidad de un cuestionario para rastreo del estrés postraumático en una población colombiana. *Revista de Neurología*, *34*(10), 911–6.
- Portela de Santana, M. L., da Costa Ribeiro Junior, H., Mora Giral, M., & Raich, R. (2012). La epidemiología y los factores de riesgo de los trastornos alimentarios en la adolescencia: una revisión. *Nutrición hospitalaria*, *27*(2), 391–401.
- Prieto, G., & Delgado, A. R. (2003). Análisis de un test mediante el modelo de Rasch. *Psicothema*, *15*(1), 94–100.
- Prieto, G., & Delgado, A. R. (2007). Measuring math anxiety (in Spanish) with the Rasch rating scale model. *Journal of applied measurement*, *8*(2), 149.
- Prieto, G., & Nieto, E. (2014). Influence of DIF on Differences in Performance of Italian and Asian Individuals on a Reading Comprehension Test of Spanish as a Foreign Language. *Journal of applied measurement*, *15*(2), 176–188.
- Prieto, G., & Velasco, A. (2003). Uso del modelo de Rasch para poner en la misma escala las puntuaciones de distintos tests. *Actualidades en psicología*, *19*(106), 5–23.
- Rasch, G. (1960). Probabilistic models for some intelligence and attainment tests. Copenhagen, Denmark: Danish Institute for Educational Research.
- Wright, B. D., & Douglas, G. A. (1976). Rasch item analysis by hand. Statistical Laboratory Department of Education Univ Chicago.

Colegio: _____ Edad: _____

Curso: _____ Sexo: _____

Marque con una (X) el espacio que corresponde a la afirmación que se ajusta más a su forma de ser.

1. Me aterroriza la idea de estar con sobrepeso

2. Me preocupo por los alimentos

3. Corto mis alimentos en trozos pequeños

4. Me preocupa el contenido calórico de los alimentos

5. Evito especialmente las comidas con alto contenido
de carbohidratos (pan, arroz, papas)

6. Me siento extremadamente culpable después haber comido

7. Me preocupo por estar más delgado(a)

8. Pienso en quemar calorías cuando hago ejercicio

9. Me preocupa la idea de tener grasa en mi cuerpo

10. Como más lento que los demás, o me demoro más
en terminar mis comidas

11. Evito los alimentos que contienen azúcar

12. Como alimentos dietéticos (diet o light)

13. Siento que la comida controla mi vida

14. Mantengo un control de lo que ingiero

15. Siento que los demás me presionan para comer

16. Me siento culpable después de comer dulces

17. Engancho con conductas de dieta

18. Me gusta sentir el estómago vacío
