

Análisis psicométrico de la Escala de Depresión CES-D bajo el modelo de crédito parcial de Rasch*

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

Recibido: 15 de junio del 2012 / Aprobado: 25 de julio del 2012

Se estudiaron las propiedades psicométricas de la Escala de Depresión del Centro de Estudios Epidemiológicos (CES-D) bajo el modelo de crédito parcial de Rasch. Se trabajó con 2400 alumnos universitarios de la ciudad de Lima, los cuales presentaron media de la edad de $20,4 \pm 5.15$ años, el 53,0% fueron mujeres y el 47,0% fueron varones. Los alumnos pertenecientes a universidades estatales fueron el 64,1% y los de universidades particulares el 35,9%. Los hallazgos indicaron que los ítems del CES-D presentan una adecuada calibración de acuerdo con el modelo de crédito parcial del modelo de Rasch. El análisis comparativo indica que no existen diferencias estadísticas significativas por sexo ni por tipo de universidad.

análisis psicométrico / depresión / CES-D / crédito parcial / Rasch

Analysis of the Depression Scale CES-D with Rasch's Model of Partial Credit

This study analyzes the psychometric characteristics of the Depression Scale of the Centro de Estudios Epidemiológicos (CES-D) utilizing the Rasch Model (partial credit) in university students in metropolitan Lima. The sample consisted of 2400 university students (mean age, 20,4 years, 53% female and 47% male, 64.1% studying at state universities and 35.9% at private universities). Results indicate that the CES-D has an appropriate calibration according to the model of Partial Credit of the Rasch Model. The comparative analysis shows no significant differences between gender and type of university.

psychometric analysis / depression / CES-D / partial credit / Rasch

* El estudio fue auspiciado por el Consejo Superior de Investigación de la Universidad Nacional Mayor de San Marcos, con el código 101801101.

Correos electrónicos: lescurra@ulima.edu.pe, aedvdt2@yahoo.com

Los métodos para medir los constructos psicológicos han variado de forma significativa desde sus inicios, debido a que han sido influenciados por la presencia de los diferentes paradigmas de investigación. En décadas pasadas imperaba el paradigma funcionalista, lo cual motivó que se tomara en cuenta la noción de control experimental, que tuvo importante influjo en el desarrollo de la estandarización y las normas o baremos de comparación. Así, se consolidó el desarrollo de la Teoría Clásica de los Tests (TCT).

En la actualidad, la presencia del paradigma cognitivo ha priorizado el concepto de proceso y su mayor influencia en el campo de la medición se ha dado sobre la base de la denominada calibración de los ítems, la función de información y el estudio del sesgo de los instrumentos, lo que ha devenido en la Teoría de la Respuesta al Ítem (TRI), especialmente en el modelo de Rasch (Hambleton, Swaminathan & Rogers, 1991; Fisher & Molenaar, 1995; Van der Linden & Hambleton, 1997; Embretson y Reise, 2000; De Ayala, 2009; Santisteban, 2009). Esto ha permitido que muchas de las pruebas psicológicas construidas inicialmente con el modelo de la TCT, sean desarrolladas con las nuevas concepciones de la TRI y el modelo de Rasch, a fin de optimizar la medición psicológica y con ello llegar a mejores decisiones.

LA ESCALA DEL CENTRO DE ESTUDIOS EPIDEMIOLÓGICOS DE LA DEPRESIÓN (CES-D)

La escala de Depresión del Centro de Estudios Epidemiológicos CES-D se creó en el Instituto Nacional de Salud Mental de los Estados Unidos (NIMH), fue desarrollada por Radloff (1977) con la finalidad de detectar síntomas depresivos en poblaciones de adolescentes y adultos. La escala permite identificar indicadores de riesgo de la depresión mediante la sintomatología. Es un instrumento para ser utilizado en comunidad, en población no clínica; su principal utilidad reside en detectar sintomatología depresiva, así como los factores de riesgo para la depresión que son de naturaleza sociodemográfica.

El instrumento ha sido traducido y validado en nuestro medio bajo el modelo de la teoría clásica de los tests por Riveros (2004), quien en el contexto de la ciudad de Lima trabajó con alumnos universitarios ingresantes. La forma de aplicación es del tipo autoadministrado. Se han desarrollado diferentes tipos de estudios con el CES-D, entre los que destacan:

- Masten, Caldwell-Colbert, Alcalá y Mijares (1986) evaluaron en una muestra mexicana la validez concurrente, la consistencia interna a través del coeficiente alfa de Cronbach y la confiabilidad test-retest del CES-D, alcanzando resultados positivos.

- Chapleski, Lamphere, Kaczynski, Lichtenberg y Dwyer (1997) aplicaron el CES-D para analizar la depresión en una muestra de 309 participantes indios americanos de los Grandes Lagos. Para ello evaluaron a nativos americanos de 55 años escogidos al azar, que vivían en zonas urbanas, rurales dentro y fuera de la reserva. Los resultados indicaron que el instrumento fue adecuado para evaluar la depresión.
- Wilcox, Field, Prodromidis y Scafidi (1998), por su parte, analizaron la relación entre el BDI y el CES-D, así como la correlación entre el BDI y el CES-D con el DISC, en una muestra de madres adolescentes, alcanzando correlaciones significativas.
- Zumbo, Gelin y Hubley (2001), por su parte, analizaron la estructura factorial y el funcionamiento diferencial del CES-D, para ello trabajaron con una muestra de 600 adultos entre 17 y 87 años de la comunidad británica. Los resultados del análisis factorial confirmatorio indicaron la existencia de una sola dimensión, observándose además que no existió un funcionamiento diferencial de acuerdo con el género.
- Reyes-Ortega, Soto-Hernández, Milla-Kegel, García-Ramírez, Hubard-Vignau, Mendoza-Sánchez, Mejía-Garza, García-Peña y Wagner Echeagaray (2003) realizaron la actualización de la Escala de Depresión del Centro de Estudios Epidemiológicos (CES-D). Para ello realizaron un estudio piloto en una muestra de 300 participantes geriátricos mexicanos. Los resultados permitieron demostrar que el instrumento presentó confiabilidad por consistencia interna, así como validez.
- Cole, Rabin, Smith & Kaufman (2004) desarrollaron y validaron una forma corta del CES-D derivado con el modelo de Rasch. Para ello evaluaron a 742 participantes, los cuales desarrollaron la forma completa del CES-D. Las respuestas fueron calibradas de acuerdo con los supuestos del modelo de Rasch.
- Aguilera, Carreño y Juárez (2004) evaluaron las características psicométricas de la CES-D en una muestra de 310 adolescentes rurales mexicanos de zonas de alta tradición migratoria. Los resultados corroboraron que el instrumento presentó validez y confiabilidad por consistencia interna.
- Yang y Jones (2007) analizaron la existencia de sesgo al responder el CES-D; para ello se aplicaron instrumentos a una muestra de 2340 participantes adultos mayores. Los resultados indicaron que 13 de los 20 ítems estaban libres de sesgo, ya sea por edad, raza o género.
- Herrero y Gracia desarrollaron una versión breve del CES-D que constaba de 7 ítems (CES-D 7); para ello trabajaron con una muestra de 1051 participantes entre 18 y 81 años de

edad. Los análisis psicométricos indicaron que la nueva versión presentó confiabilidad por consistencia interna y por test-retest y validez, además los análisis reportaron que se obtenían resultados equivalentes a la versión completa.

- Tiburcio y Natera (2007) adaptaron el CES-D al contexto indígena Ñahñú de México, para lo cual trabajaron con una muestra de 191 mujeres, los resultados permitieron corroborar que el instrumento presentó una adecuada confiabilidad por consistencia interna y validez.
- Stahl, Sum, Lum Liow, Chan, Verma, Chua y Chong (2008) evaluaron la confiabilidad y la validez del CES-D en una muestra multiétnica que se atendían por diabetes en la ciudad de Singapur. Para ello trabajaron con una muestra de 291 casos. Los resultados permitieron identificar que el CES-D presentaba buenas cualidades psicométricas en la consistencia interna.
- Finalmente, Tatar & Saltukoglu (2010) realizaron la adaptación del CES-D al idioma turco; trabajaron con una muestra de 1143 participantes y, de acuerdo con la TCT y la TRI, los hallazgos indicaron que el instrumento presentó confiabilidad por consistencia interna, así como confiabilidad por estabilidad con el test-retest. También presentó una adecuada calibración de los ítems

con el modelo de 2 parámetros de la TRI y adicionalmente se estudió la validez de constructo aplicando análisis factoriales exploratorios y confirmatorios, los cuales corroboraron la existencia de 4 factores.

EL MODELO DE CRÉDITO PARCIAL (MCP)

El modelo de crédito parcial es un modelo psicométrico que pertenece a la familia de modelos de Rasch. Este permite analizar ítems que corresponden a respuestas graduadas en varios niveles de cuantificación, como es el caso de la escala tipo Likert (Nering & Ostini, 2010; De Ayala, 2009; Ostini & Nering, 2006; Martínez, Hernández & Hernández, 2006; Embretson & Reise, 2000; Van der Linden & Hambleton, 1997; Fischer & Molenaar, 1995; Masters, 1982).

El modelo estima la probabilidad de emitir una respuesta de cada nivel de cuantificación de acuerdo con la ecuación:

$$p_{jk}(\theta_i) = \frac{\exp \left[\sum_{j=0}^k (\theta_i - \beta_{jk}) \right]}{\sum_{r=0}^{m_j} \left[\exp \sum_{j=0}^r (\theta_i - \beta_{jk}) \right]}, \quad k = 1, \dots, m_j$$

El modelo permite calcular la probabilidad de responder en una cierta categoría b_{jk} que representa la dificultad del paso k (mientras más alto el valor, más difícil o complejo de realizar) que es un paso en particular dentro del ítem.

La ecuación establece que la probabilidad de dar una respuesta de un rendimiento/calidad determinada ($X=c$) para un sujeto, dado su nivel de habilidad (q) y el conjunto de dificultades asociadas a cada paso del ítem ($x=d_{11}, d_{12}, d_{13}$), corresponde a la diferencia entre la habilidad del sujeto y la dificultad del paso asociado a cada categoría, considerando todos los pasos del ítem en conjunto. Un ítem de c categorías tendrá un número de pasos $m=c-1$.

El MCP y los modelos de Rasch, en general, son aplicados en la psicometría contemporánea debido a que presentan propiedades relevantes para la construcción y análisis de pruebas psicológicas en general, entre las que destacan:

- a) Presentan estadísticos suficientes, lo cual implica que cuando los datos se ajustan al modelo, la puntuación directa total (como es el caso de la suma de los puntajes de las respuestas) presenta toda la información necesaria para estimar la habilidad de los sujetos.
- b) Capacidad de producir medidas sumativas, las cuales presentan propiedades equivalentes a las escalas de intervalo.
- c) Especificidad objetiva, la cual indica que los parámetros de los ítems son independientes de la habilidad de los sujetos con los cuales se les estima y a su vez que los parámetros de la habilidad de la persona son independientes de los ítems.

El MCP ha sido utilizado en el campo de la investigación psicológica utilizando diferentes procedimientos y estrategias (Rojas Gonzales, Padilla & Pérez, 2000). Entre los estudios llevados a cabo destacan:

- Smith, Rush, Fallowfield, Velikova & Sharpe (2008) desarrollaron un estudio en pacientes con cáncer, tomando como instrumento el HADS, analizando los estadísticos de la forma de Rasch y el tamaño de la muestra para datos politómicos. Los resultados demostraron que los estadísticos eran sensibles al tamaño de la muestra.
- Covic, Pallant, Tennant, Cox, Emery, & Conaghan (2009) estudiaron la variabilidad en la depresión y su prevalencia en pacientes iniciales con artritis reumatoide comparados a través del CES-D y el HAD-D; para ello trabajaron con 92 pacientes, los cuales respondieron a los 2 instrumentos. Los resultados permitieron concluir que el CES-D fue más consistente para ser usado con dichos pacientes, pero necesitaba ser reducida a 13 ítems. Mientras que el HADS-D era válido para los pacientes y además se encontró que las dos escalas presentaban de forma subyacente el mismo constructo, pero los puntos de cortes permiten estimar diferentes niveles de depresión, aunque el CES-D detecta una mayor prevalencia de la depresión que el HADS-D.

- Shea, Tennant & Pallant (2009) evaluaron con el MCP de Rasch la escala de depresión, ansiedad y estrés (DASS). Para ello trabajaron con una muestra de 745 participantes. Los resultados indicaron la necesidad de remover un ítem de cada una de las subescalas del DASS para alcanzar un rendimiento adecuado en el modelo. Las escalas reducidas mostraron una adecuada consistencia interna, confiabilidad, unidimensionalidad y libertad en el rendimiento del funcionamiento diferencial por sexo, edad y modo de distrés psicológico en general. La escala que combina la ansiedad y el estrés mostró una satisfactoria forma en el modelo de Rasch después de eliminar 3 ítems.
- Forkmann, Boecker, Wirtz, Glaesmer, Brähler, Norra & Gauggel (2010) validaron la escala de detección de depresión basada en el modelo de Rasch (DESC), para una muestra general alemana. Para ello trataron de examinar el modelo de Rasch y la validez convergente y discriminante trabajando en una muestra de 2509 casos.

OBJETIVOS

- 1) Analizar las propiedades psicométricas de los ítems que conforman la escala CES-D, de acuerdo con la teoría clásica de los tests.
- 2) Evaluar el ajuste psicométrico de los ítems que conforman la escala

CES-D, de acuerdo con la teoría de la respuesta al ítem según el modelo de crédito parcial de Rasch.

- 3) Analizar la función de información que presenta el instrumento.
- 4) Evaluar la validez de constructo de la escala CES-D a través del análisis de los componentes principales de los residuos Rasch.
- 5) Determinar la confiabilidad de las medidas de la escala CES-D.

MÉTODO

Diseño de investigación

Para el desarrollo del estudio se utilizó un diseño de investigación descriptivo y psicométrico (Sánchez & Reyes, 2006; Alarcón, 2008).

Universo y muestra de investigación

El universo estuvo conformado por alumnos universitarios entre primer y tercer año de estudio, matriculados en universidades nacionales y particulares de la ciudad de Lima. Se trabajó con una muestra representativa de 2400 casos. Para la obtención de la muestra se aplicó un procedimiento de muestreo probabilístico estratificado de acuerdo con el tamaño de las universidades seleccionadas y el tamaño de las facultades.

La media de la edad de los alumnos evaluados fue de $20,4 \pm 5.15$ años, el 53,0% fueron mujeres y el 47,0% fue-

ron varones. Los alumnos pertenecientes a universidades estatales constituyeron el 64,1% y los de particulares el 35,9%.

Variable de estudio

La variable estudiada corresponde a la medición de la depresión mediante la escala CES-D. Adicionalmente, los resultados fueron comparados según el tipo de universidad de procedencia y el género de los alumnos.

Instrumentos de recolección de datos

Se utilizó la escala CES-D, que comprende 20 ítems relacionados con el humor, el comportamiento y la percepción, que fueron considerados relevantes en estudios clínicos sobre depresión (Radloff, 1977). Fue construida siguiendo los criterios para la elaboración de escalas tipo Likert con enunciados y niveles de calificación. Consta de 20 reactivos representativos de los componentes más importantes de la sintomatología depresiva, y, de acuerdo con el análisis factorial de la escala, los reactivos quedan agrupados en cuatro factores: a) afecto depresivo; b) afecto positivo; c) actividad somática; y d) relaciones interpersonales.

La escala de calificación implica indicar la respuesta de la siguiente manera, su aplicación consiste en marcar con una "X" la respuesta que mejor describa cómo se ha sentido la persona durante la semana pasada, la semana previa a la

aplicación y las respuestas se califican de la siguiente manera: 0 = menos de un día; 1 = uno o dos días; 2 = tres o cuatro días, y 3 = cinco o siete días. El puntaje total de la escala es la sumatoria de los puntos de los veinte (20) reactivos, que tienen un rango de 0 a 60.

Radloff (1977) reporta que la consistencia interna original evaluada a través del coeficiente Alfa de Cronbach fue de 0.85 en la muestra general y de 0.90 en la muestra de pacientes.

Para el desarrollo del estudio se utilizó la escala CES-D de Radloff (1977) traducida y adaptada por Riveros (2004) en estudiantes universitarios de la ciudad de Lima, quien también demostró que tenía consistencia interna a través del coeficiente Alfa de Cronbach, así como validez de constructo.

Procedimiento de recolección de datos

Para obtener los datos se trabajó de la siguiente manera: a) en primer lugar se estableció para cada universidad y facultad participante la proporción de casos a ser evaluados; b) de acuerdo con los resultados obtenidos, y siguiendo un procedimiento de muestreo estratificado, se eligieron de forma aleatoria las aulas de las facultades; c) una vez seleccionadas las aulas que debían ser evaluadas, se procedió a realizar las coordinaciones necesarias para aplicar el instrumento. La aplicación se realizó de forma grupal y se tomaron en cuenta

los alumnos presentes en el momento de la evaluación.

Técnicas de procesamiento y análisis de datos

El análisis estadístico se realizó tomando en cuenta las recomendaciones propuestas por Muñiz (1990, 1996, 1997), quien incluyó los siguientes análisis:

- a. Análisis psicométrico de la calibración de la escala CES-D. Se realizó con el programa Winstep (Linacre, 2007).
- b. Análisis descriptivo, que se utilizó para analizar la forma como se distribuyeron los puntajes del test; para ello se utilizó la prueba de bondad de ajuste a la curva normal de Kolmogorov-Smirnov.

Tabla 1
Análisis psicométrico con la TCT del CES-D

Ítem	M	D. E.	ritc
1	1.06	0.96	0.64
2	0.92	0.97	0.65
3	1.19	1.00	0.79
4	1.06	0.84	0.61
5	1.49	0.81	0.56
6	1.32	0.96	0.78
7	1.48	0.84	0.35
8	1.10	0.93	0.73
9	0.85	0.97	0.84
10	1.15	0.91	0.72
11	1.10	1.02	0.66
12	1.10	1.05	0.83
13	1.05	0.83	0.60
14	1.35	0.97	0.76
15	0.86	0.84	0.61
16	1.10	0.98	0.78
17	0.83	0.90	0.72
18	1.35	0.96	0.82
19	1.02	0.94	0.73
20	1.06	0.96	0.64

Alfa de Cronbach = 0.95
Intervalos de confianza del Alfa de Cronbach = 0.91 – 0.98

N = 2400

Tabla 2
Medida, error y ajuste de los ítems del CES-D

Ítem	Medida	E. E.	Infit	Outfit	rpbis
1	0.17	0.04	1.23	1.23	0.68
2	0.53	0.04	1.25	1.31	0.68
3	-0.15	0.04	0.83	0.79	0.81
4	0.17	0.04	1.06	1.11	0.64
5	-0.85	0.04	1.06	1.09	0.61
6	-0.45	0.04	0.76	0.75	0.81
7	-0.83	0.04	1.28	1.24	0.40
8	0.07	0.04	0.89	0.88	0.75
9	0.71	0.04	0.78	0.71	0.83
10	-0.05	0.04	0.88	0.88	0.75
11	0.08	0.04	1.29	1.26	0.70
12	0.06	0.04	0.83	0.80	0.84
13	0.19	0.04	1.07	1.11	0.64
14	-0.51	0.04	0.82	0.80	0.79
15	0.69	0.04	1.16	1.15	0.80
16	0.07	0.04	0.83	0.85	0.74
17	0.77	0.04	0.96	0.90	0.83
18	-0.51	0.04	0.65	0.65	0.84
19	0.26	0.04	0.93	0.91	0.76
20	-0.43	0.04	0.92	0.93	0.74

N = 2400

- c. El análisis comparativo de los puntajes se realizó con base en la prueba U de Mann-Whitney, a fin de evaluar la existencia de diferencias por tipo de universidad y género.

RESULTADOS

Análisis psicométrico con la TCT

El análisis psicométrico del CES-D con la teoría clásica de los tests (TCT), incluido en la tabla 1 permite apreciar que

los ítems alcanzan correlaciones ítem-test que fluctúan entre 0.35 (ítem 7) y 0.84 (ítem 9), es decir que todos superan el criterio de ser mayores de 0.20 (Kline, 1993). Adicionalmente, en análisis de la confiabilidad por consistencia interna a través del coeficiente Alfa de Cronbach presenta un valor de 0.95 y un intervalo de confianza entre 0.91 y 0.98, por lo que se concluye que el instrumento permite obtener puntajes confiables. Este hallazgo corrobora que el CES-D es adecuado con la TCT.

Análisis psicométrico con el MCP

Los resultados del análisis psicométrico de la calibración de los ítems del CES-D con el modelo de crédito parcial de Rasch (véase la tabla 2), permiten apreciar que los parámetros del ítem de medida presentan valores que varían entre los niveles bajos (-0.85), como es el caso del ítem 5, y los niveles altos (0.77), como el ítem 17, es decir que existe un espectro intermedio de los niveles evaluados en el constructo. Adicionalmente, se observa que en los ítems el error de estimación (E.E.) es pequeño. La revisión de los errores Infit, que evalúan el ajuste interno de los ítems, y de los Outfit, que evalúan el ajuste externo de estos, permite revelar que todos los ítems cumplen con

los criterios de inclusión; con el modelo de Rasch. Los resultados permiten concluir que la hipótesis 1 es válida.

En relación con la revisión del test de la Función de Información de los ítems del CES-D (véase la figura 1), se puede notar que las mediciones más estables se pueden obtener en el caso de que la habilidad del sujeto fluctúe entre -3 logit y 3 logit, de tal manera que cuando la habilidad presenta valores diferentes de los indicados, ya sean más bajos o más altos, el resultado alcanzado es más impreciso y presenta más error, especialmente en los extremos.

Para estudiar la validez, se obtuvo evidencias sobre la validez de constructo del CES-D, mediante el análisis de la unidimensionalidad de los ítems, lo

Figura 1
Función de Información del CES-D

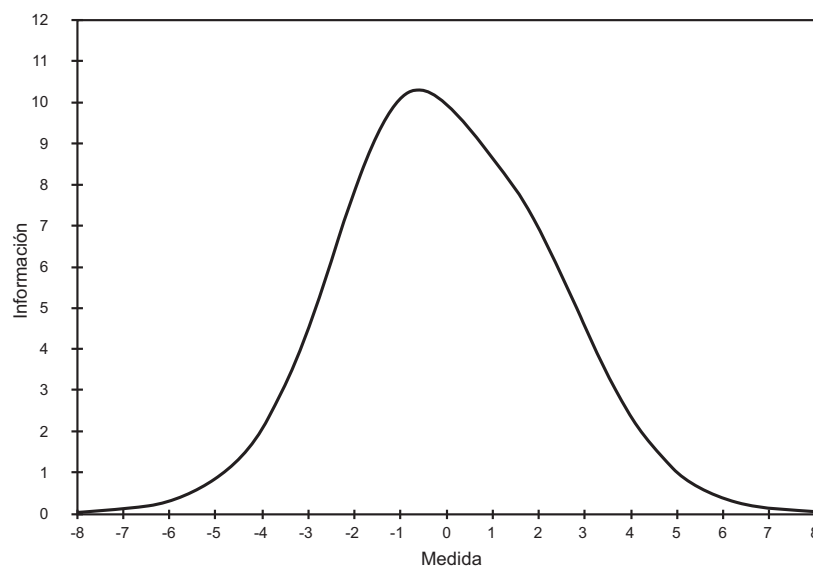


Tabla 3
Varianza estandarizada residual (en unidades de *eigenvalues*)

Fuentes	Autovalor	Empírica (%)	Modelada (%)
Total de varianza en observaciones =	57.5	100,0	100,0
Varianza explicada por las medidas =	47.5	82,6	82,6
Varianza no explicada (total) =	10.0	17,4	17,4
Var. no explicada en el primer contraste =	2.3	4,0	
Var. no explicada en el segundo contraste =	1.6	2,8	
Var. no explicada en el tercer contraste =	1.6	2,7	
Var. no explicada en el cuarto contraste =	1.4	2,5	
Var. no explicada en el quinto contraste =	1.3	2,2	

N = 2400

cual implica que el instrumento mide un solo constructo o rasgo latente (Hattie, 1985).

Un instrumento es unidimensional si las respuestas dadas a él son producidas en base a un único atributo. Wright y Linacre (1998) indican que ningún instrumento puede ser perfectamente unidimensional y lo que se busca es tener instrumentos que en esencia muestren unidimensionalidad.

En este caso, se realizó un análisis de componentes principales de los residuos (ACPR), luego de ajustar el modelo de Rasch unidimensional de acuerdo con lo que propone Linacre (2007), quien indica que si el primer autovalor contiene menos del 5% de varianza o si su valor es menor que 3.0 no existirían serias evidencias que atenten contra el supuesto de unidimensionalidad y, por lo tanto, los datos se pueden analizar adecuadamente utilizando los modelos de Rasch. Estas condiciones se cumplen

tal como se muestra en la tabla 3, por lo que se concluye que se está midiendo un constructo unidimensional con el CES-D. En conclusión, los análisis realizados demuestran que la escala CES-D se ajusta bien a un modelo Rasch unidimensional.

Análisis de la confiabilidad de las medidas

Las evidencias sobre la confiabilidad de las medidas fueron establecidas de acuerdo con el índice de separación de personas. Este índice indica qué tan bien sirven las medidas de un test para diferenciar las cantidades de rasgo latente que poseen los evaluados (Wright & Masters, 1982). Un índice menor a 0,50 indica que las diferencias entre las medidas son producidas principalmente por el error de medición (Fisher, 1992). Como se aprecia en la tabla 4, los resultados indican que los valores obtenidos fueron elevados.

Tabla 4
Confiabilidad de las medidas derivadas de la aplicación del CES-D

Estadístico	Real	Modelo
Pearson confiabilidad	0.93	0.94
Índice de separación de personas	3.75	4.09
RMSE	0.40	0.37

N = 2400

Análisis descriptivo

El análisis de la bondad de ajuste a la curva normal realizado a través de la prueba de Kolmogorov-Smirnov (véase la tabla 5) permite notar que los puntajes del CES-D ($M = 22.78$ $DE = 13.67$) presentan un estadístico que es significativo ($Z = 0.15$ $P < .000$), por lo cual se puede concluir que no se aproxima a la distribución normal.

El análisis por universidad de procedencia y por género también permite notar que existen diferencias estadísticas significativas en todos los casos, por lo que se concluye que por las variables de comparación no existen adecuadas

aproximaciones a la distribución normal, de manera que se aplicaron pruebas estadísticas no paramétricas para ejecutar los análisis comparativos.

Análisis comparativo

El análisis comparativo de los puntajes del CES-D según el tipo de universidad de procedencia, tal como se presenta en la tabla 6, permite apreciar que no existen diferencias estadísticas significativas ($Z = -0.70$ $p > .05$), lo cual indica que los alumnos presentan valores similares en la escala de depresión.

Respecto al análisis comparativo de los puntajes del CES-D según el género

Tabla 5
Análisis de la bondad de ajuste a la curva normal a través de la prueba de Kolmogorov-Smirnov de las variables analizadas

Variable	n	M	D. E.	KS-Z	Sig.
U. estatal	1538	22.94	13.71	0.10	.000
U. particular	862	22.51	13.61	0.11	.000
Varón	1128	22.65	13.72	0.13	.000
Mujer	1272	22.90	13.63	0.14	.000
Total	2400	22.78	13.67	0.15	.000

Tabla 6
Análisis comparativo del CES-D por tipo de universidad

Universidad de procedencia	Estatal	Particular	U de Mann-Whitney	Z	Sig.
	N = 1538	N = 862			
	MR	MR			
CES-D	1207.93	1187.24	651448.50	-0.70	.483

Tabla 7
Análisis comparativo del CES-D por género

Universidad de procedencia	Varón	Mujer	U de Mann-Whitney	Z	Sig.
	N = 1128	N = 1272			
	MR	MR			
CES-D	1192.42	1207.67	708293.50	-0.54	.590

de los participantes (véase la tabla 7), permite notar que no existen diferencias estadísticas significativas ($Z = -0.54$ $p > .05$), lo cual indica que los alumnos por género presentan valores similares en la escala de depresión.

DISCUSIÓN

El análisis de los resultados de la escala CES-D permite apreciar que las correlaciones ítem-test son adecuadas y presenta una confiabilidad por consistencia elevada, resultado consistente con los estudios realizados de Tatar & Saltukoglu (2010); Stahl et al. (2008), Tiburcio & Natera (2007); Aguilera, Carreño & Juárez (2004); Riveros

(2004); Reyes-Ortega et al. (2003); Zumbo, Gelin & Hubley (2001); Masten et al. (1986), de manera que cumple con los requisitos psicométricos de acuerdo con los supuestos de la teoría clásica de los test (Kline, 1993; Muñiz, 1996; Anastasi & Urbina, 1998).

En lo que corresponde al análisis de acuerdo con los supuestos del modelo de crédito parcial de Rasch, los resultados indican que los hallazgos son consistentes, lo que podría deberse a que el instrumento estaría evaluando de forma prioritaria el constructo de la Depresión, motivo por el cual cumplen con los requerimientos psicométricos básicos para este tipo de instrumento en el

modelo utilizado (De Ayala, 2009; Furr & Bacharach, 2008; Martínez, Hernández & Hernández, 2006; Muñiz, 1996; Anastasi & Urbina, 1998).

Los hallazgos obtenidos con relación al análisis gráfico permite identificar que los ítems cubren de manera adecuada el espectro del constructo que se evalúa, este es un resultado que refleja la posibilidad de obtener valoraciones que cumplan un rango amplio de las actividades que definen a los constructos analizados, tal como lo proponen: De Ayala (2009), Embretson & Reise (2000), Van Der Linden & Hambleton, (1997).

En relación con los análisis de las funciones de información los hallazgos son positivos, pues también el rango de los valores con puntajes precisos presenta una amplitud adecuada tal como lo proponen De Ayala (2009); Muñiz (1997, 1996); y Hambleton, Swaminathan & Rogers (1991); además, puede notarse que los puntajes con mayores errores e imprecisiones se presentan en los valores de los extremos inferiores y superiores de la distribución de puntajes.

Otro resultado importante corresponde al análisis de la unidimensionalidad propuesto por Linacre (2007), debido a que la metodología del MCP aplicada al instrumento permite identificar la existencia de una sola dimensión subyacente. Este hallazgo concuerda con el estudio desarrollado por Zumbo et al. (2001), aunque se contraponen con lo reportado por Tatar

& Saltukoglu (2010) y Riveros (2004), quienes proponen la existencia de más de un factor, lo cual podría deberse a la diferente metodología empleada para el estudio de los factores, ya que dichos autores aplicaron análisis factoriales exploratorios, mientras que en el estudio se aplicó la metodología de Wright & Linacre (1998) y Linacre (2007).

Los análisis de la bondad de ajuste a la curva normal realizados a través de la prueba de Kolmogorov-Smirnov (Siegel & Castellan, 1995) permitieron concluir que los puntajes no presentan una tendencia a parecerse a la distribución normal. Esto quiere decir que el puntaje total presenta un sesgo negativo en el que predominan los valores inferiores a la media; por ello, es necesario analizar los datos con estadísticas no-paramétricas a fin de alcanzar resultados relevantes.

En general, los resultados obtenidos coinciden con los hallazgos reportados por Cole, Rabin, Smith & Kaufman (2004), lo cual corroboraría que el instrumento se adecua a la teoría de respuesta al ítem.

De los hallazgos reportados se deriva la imperiosa necesidad de realizar estudios que permitan adaptar las pruebas psicológicas que se usan de forma profesional por los psicólogos de nuestro medio, haciendo uso de metodologías psicométricas como la de Rasch, de manera que obtengan mediciones más precisas y de calidad.

CONCLUSIONES

- a. Los ítems de la escala CES-D presentan una adecuada calibración de acuerdo con el modelo de crédito parcial de Rasch.
- b. La escala CES-D presenta un adecuado ajuste al modelo de crédito parcial de Rasch de la teoría de la respuesta al Ítem.
- c. La escala CES-D presenta una adecuada función de información.
- d. Los alumnos de universidades estatales y particulares presentan puntajes similares en la escala CES-D.
- e. Las alumnas y los alumnos universitarios presentan puntajes similares en la escala CES-D.

RECOMENDACIONES

A partir de los resultados, la discusión y las conclusiones se plantean las siguientes recomendaciones:

- Desarrollar nuevas investigaciones con el CES-D revisado con el modelo de crédito parcial a fin de establecer mayores evidencias de validez.
- Desarrollar investigaciones psicométricas similares a la realizada, pero considerando muestras diferenciadas a fin de establecer la generalidad de los resultados encontrados.
- Aplicar el modelo de crédito parcial para analizar las propiedades psicométricas de otros instrumentos psicológicos utilizados en nuestro con-

texto, que correspondan a las escalas tipo Likert.

REFERENCIAS

- Aguilera, R. M., Carreño, M. S., & Juárez, F. (2004). Características psicométricas de la CES-D en una muestra de adolescentes rurales mexicanos de zonas con alta tradición migratoria. *Salud Mental, 27*(6), 57-66.
- Alarcón, R. (2008). *Métodos y diseños de investigación del comportamiento*. Lima: Editorial Universitaria.
- Anastasi, A., & Urbina, S. (1998). *Test psicológicos*. México: Prentice Hall.
- Brown, F. (1980). *Principios de la medición en psicología y educación*. México: El Manual Moderno.
- Chapleski, J. K., Lamphere, R., Kaczynski, P. A., Lichtenberg, P. A., & Dwyer, J. W. (1997). Structure of a depression measure among American Indian elders: Confirmatory factor analysis of the CES-D. *Research on Aging, 19*, 462-485.
- Cole, J., Rabin, A., Smith, T., & Kaufman, A. (2004). Development and validation of a Rasch Derived CES-D Short Form. *Psychological Assessment, 16*(4), 360-372. DOI: 10.1037/1040-3590.16.4.360.
- Covic, T., Pallant, J., Tennant, A., Cox, S., Emery, P., & Conaghan, P. (2009). Variability in depression prevalence in early rheumatoid arthritis: a com-

- parison of the CES-D and HAD-D Scales. *BCM Musculoskeletal Disorders*, 10, 18, DOI:10.1186/1471-2474-10-18.
- De Ayala, R. J. (2009). *The theory and practice of item Response Theory*. Nueva York: Guilford Publications Inc.
- Embretson, S., & Reise, S. (2000). *Item Response Theory for psychologists*. Nueva Jersey: Lawrence Erlbaum Associates Publishers.
- Fisher, W. (1996). Reliability statistics. En J. Linacre (ed.). *Rasch Measurement Transactions*. Part 2. Chicago: MESA Press.
- Fisher, G. H., & Molenaar, I. (1995). *Rasch Models, foundations, recent developments and applications*. Nueva York: Springer Verlag.
- Forkmann, T., Boecker, M., Wirtz, M., Glaesmer, H., Brähler, E., Norra, C., & Gauggel, S. (2010). Validation of the Rasch-based depression screening in a large scale German general population sample. *Health and Quality of Life Outcomes*, 8, 105.
- Furr, M., & Bacharach, V. (2008). *Psychometrics an introduction*. Los Angeles: Sage Publications.
- Hambleton, R., Swaminathan, H., & Rogers, J. (1991). *Fundamentals of item Response Theory*. Londres: Sage Publications Ltd.
- Hattie, J. (1985). Methodology review: Assessing unidimensionality of tests and items. *Applied Psychological Measurement*, 9(2), 139-164.
- Herrero, J., & Gracia, E. (2007). Una medida breve de la sintomatología depresiva (CESD-7). *Salud Mental*, 30(5), 40-46.
- Kline, P. (1993). *The handbook of psychological testing*. Londres: Routledge.
- Linacre, J. M. (2007). *A user's guide and manual to winstep, Rasch model computer programs*. Recuperado el 14 de febrero de 2011, de www.winsteps.com.
- Martínez, M. R., Hernández, M. J., & Hernández, M. V. (2006). *Psicometría*. Madrid: Alianza Editorial.
- Masten, W. G., Caldwell-Colbert, A. T., Alcala, S. J., & Mijares, B. E. (1986). Confiabilidad y validez de la escala de depresión del Centro de Estudios Epidemiológicos. *Hispanic Journal of Behavioral Sciences*, 8(1), 77-84.
- Masten, W. G., Caldwell-Colbert, A. T., William, V., Jerome, W., Mosby, L., Barrios, Y., & Helton, J. (2003). Diferencias de género en síntomas depresivos entre adolescentes mexicanos. *Anales de Psicología* 19(1), 91-95.
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149-174.
- Muñiz, J. (1990). *Teoría de respuesta a los ítems: un nuevo enfoque en la*

- evolución psicológica y educativa*. Madrid: Pirámide.
- Muñiz, J. (1996). *Psicometría*. Madrid: Universitas.
- Muñiz, J. (1997). *Introducción a la teoría de respuesta a los ítems*. Madrid: Pirámide.
- Nering, M., & Ostini, R. (2010). *Handbook of polytomous response theory models*. Nueva York: Routledge Academic.
- Ostini, R., & Nering, M. (2006). *Polytomous response theory models*. Londres: Sage Publications
- Radloff, L. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 1*, 385-401.
- Reyes-Ortega, M., Soto-Hernández, A. L., Milla-Kegel, J. G., García-Ramírez, A., Hubard-Vignau, L., Mendoza-Sánchez, H., Mejía-Garza, L. A., García-Peña, M. C., & Wagner-Echeagaray, F. A. (2003). Actualización de la Escala de Depresión del Centro de Estudios Epidemiológicos (CES-D). Estudio piloto en una muestra geriátrica mexicana. *Salud Mental, 26*(1), 59-68.
- Riveros, M. (2004). *Características sociodemográficas y niveles de depresión en estudiantes ingresantes a la UNMSM, a través de la escala CES-D*. (Tesis de maestría en Psicología con mención en Psicología clínica). Lima: Universidad Nacional Mayor de San Marcos, Facultad de Psicología, Unidad de Posgrado.
- Rojas, A., González, A., Padilla, J., & Pérez, C. (2000). Comparación de estrategias de ajuste de los datos al modelo de crédito parcial. *Psicothema, 12*(2), 296-302.
- Sánchez, C. H., & Reyes, M. C. (2006). *Metodología y diseños de la investigación científica*. Lima: Visión Universitaria.
- Santisteban, C. (2009). *Principios de psicometría*. Madrid: Síntesis.
- Shea, T., Tennant, A., & Pallant, J. (2009). Rasch model analysis of the depression, anxiety and stress scale (DASS). *BCM medical research methodology, 9*, 21. DOI:10.1186/1471-244X-9-21.
- Siegel, S., & Castellan, N. (1995). *Estadística no paramétrica: Aplicada a las ciencias de la conducta*. México: Trillas.
- Smith, A., Rush, R., Fallowfield, L., Velikova, G., & Sharpe, M. (2008). Rasch fit and sample size considerations for polytomous data. *BCM medical research methodology, 8*, 33. DOI:10.1186/1471-2288-8-33.
- Stahl, D., Sum, C. F., Lum, S. S., Liow, P. H., Chan, Y. H., Verma, S., Chua, H. C., & Chong, S. A. (2008). Screening for depressive symptoms: Validation of the CES-D scale in a multi-ethnic group of patient with diabetes in Singapore. *Diabetes Care, 31*(6), 1118-1119.

- Tatar, A., & Saltukoglu, G. (2010). The adaptation of the CES-Depression Scale into Turkish through the use of confirmatory Factor Analysis and item response Theory and the examination of psychometric characteristics. *Bulletin of Clinical Psychopharmacology*, 20, 213-227.
- Tiburcio, M., & Natera, G. (2007). Adaptación al contexto Ñahñú del cuestionario de enfrentamientos (CQ), la escala de síntomas (SRT) y la escala de depresión del Centro de Estudios Epidemiológicos (CES-D). *Salud Mental*, 30(3), 18-58.
- Van der Linden, W., & Hambleton, R. (1997). *Handbook of Modern Item Response Theory*. Nueva York: Springer Verlag.
- Walsh, W., & Betz, Nancy (1995). *Tests and assessment*. Nueva Jersey: Prentice Hall.
- Wilcox, H., Field, T., Prodromidis, M., & Scafidi, F. (1998). Correlations between the BDI and CES-D in a sample of adolescent mothers. *Adolescence*, 33(131), 565-74.
- Wright, B., & Linacre, J. (1998). MESA research memorandum 44. *Archives of Physical Medicine and Rehabilitation*, 70(12), 857-860.
- Yang, F. M., & Jones, R. N. (2007). Center for Epidemiologic Studies-Depression Scale (CES-D), item Response Bias found with Mantel-Haenszel Method successfully replicated using latent variable modeling. *Journal of Clinical Epidemiology*, 60(11), 1195-1200.
- Zumbo, B. D., Gelin, M. N., & Hubley, A. M. (2001). Psychometric study of the CES-D: Factor analysis and DIF. *Journal of Clinical Epidemiology*, 54, 343-349.