



---

# EVALUACIÓN DEL FUNCIONAMIENTO DIFERENCIAL DE LOS ÍTEMS (DIF) DE LA ESCALA DE DEPRESIÓN DEL CENTRO PARA ESTUDIOS EPIDEMIOLÓGICOS (CES-D) RESPECTO DEL GÉNERO EN UNA MUESTRA DE ADOLESCENTES MEXICANOS.

Jesús Rafael Osorno Munguía, Alma Vallejo Casarín, Beatriz Segura Celis-Ochoa  
y Teresa de Jesús Mazadiego Infante<sup>1</sup>.

Universidad Veracruzana  
Zona Poza Rica-Tuxpan.

## Resumen

El objetivo de este trabajo fue analizar la presencia de funcionamiento diferencial de los ítems (DIF, *Differential Ítem Functioning* por sus siglas en inglés) respecto del género en la CES-D en una muestra de adolescentes. La evaluación se llevó a cabo con 1011 adolescentes mexicanos en edades de 11 a 21 años, 486 mujeres y 525 varones, estudiantes de escuelas secundarias y preparatorias públicas del Estado de Veracruz, México, de familias pertenecientes a un nivel socioeconómico bajo y medio. Se realizó el análisis del DIF mediante los métodos de Mantel-Haenszel (Holland y Thayer, 1988) y de Regresión Logística (Swaminathan y Rogers, 1990) en sus versiones iterativas. Se calculó la concordancia entre los métodos y se identificaron ocho de veinte ítems, con ambos métodos con presencia de DIF significativo. Los resultados nos sugieren una revisión del CES-D en cuanto al género, se discute la presencia de calificaciones altas en la depresión en las mujeres.

**Palabras Clave:** Funcionamiento diferencial de los Ítems, CES-D, Sesgo de género, Adolescencia, Evaluación psicométrica.

---

<sup>1</sup> Profesores de Tiempo Completo de la Facultad de Psicología de la Universidad Veracruzana, Región Poza Rica Tuxpan. Miembros del Cuerpo Académico (CA): Educación Salud y Evaluación. Este trabajo fue financiado con fondos del PIFI para el CA. Blvd., Lázaro Cárdenas 801, Col Morelos, Poza Rica Veracruz, cp. 93340, e-mail: [josorno@uv.mx](mailto:josorno@uv.mx), [avallejo@uv.mx](mailto:avallejo@uv.mx)

### Abstract

The aim of this study was to analyze the presence of differential item functioning (DIF) for the gender in the CES-D in a sample of adolescents. The evaluation was carried out with 1011 Mexican adolescents, aged 11 to 21 years, 486 women and 525 men, students from public junior high and high schools of the state of Veracruz, México, families belonging to a low and a half socioeconomic status. The analysis was performed using DIF methods Mantel-Haenszel (Thayer and Holland, 1988) and Logistic Regression (Swaminathan and Rogers, 1990) in their iterative versions. The correlation between the methods was calculated and identified eight of the total of twenty items with both methods with significant presence of DIF. The results suggest a revision of the CES-D in terms of gender, the presence of high ratings on depression in women is discussed.

**Key words:** Differential item functioning, CES-D, Gender bias, Adolescence, Psychometric analysis.

### Introducción

La adolescencia es reconocida como una etapa de cambios importantes en diferentes ámbitos el fisiológico, afectivo, familiar y escolar (Estévez, Musitu y Herrero, 2005), estos cambios implican altos grados de estrés que el individuo debe enfrentar (Washburn, Hillman & Sawiloswsky, 2004). Es reconocido por los psicólogos investigadores y clínicos que la forma cómo el individuo enfrente el estrés tendrá honda repercusión en el ajuste y salud mental del individuo, en específico, se ha identificado una fuerte relación entre el estrés de los adolescentes y la depresión (Barra, Cerna, Kramm y Véliz, 2006).

La depresión es llamada la enfermedad del siglo XXI, se estima que en 2020 será la segunda causa de años de vida saludable perdidos a escala mundial y la primera en países desarrollados (Belló, Puentes, Medina y Lozano, 2005). En los Estados Unidos existe evidencia que indica diferencias de género (a favor de las mujeres), en los síntomas depresivos entre los adolescentes. Un estudio epidemiológico que involucró individuos de Estados Unidos, Puerto Rico y otros cinco países indicó un incremento de la depresión en la adolescencia. Además las mujeres fueron 66% más tendientes a la depresión (Masten, Caldwell-Colbert, Williams, Jerome, Mosby, Barrios, y Helton, 2003). Investigaciones anteriores han indicado un alto nivel de depresión entre las adolescentes comparadas con los

adolescentes y las explicaciones de esta diferencia se ha explorado a profundidad por científicos como Allgood-Merton, Lewinsohn y Hops (1990), que encontraron relación con la imagen corporal y la autoestima; Katragadda y Tidwell (1998) encontraron que el género fue un fuerte predictor de la depresión, así como la autoestima y los altos niveles de estrés; en este mismo estudio se exploró la depresión entre diferentes grupos étnicos de adolescentes: anglos, africanos, mexicanos e hispano americanos todos viviendo en Estados Unidos, encontrando que las mujeres hispanas tuvieron el más alto porcentaje (31.2%) del nivel de depresión moderada a severa, demostrando así que las mujeres manifiestan más síntomas depresivos que los varones en todos los grupos estudiados.

La depresión en México ocupa un lugar prominente entre los trastornos afectivos (el 53% de éstos), de acuerdo con la Encuesta Nacional de Epidemiología Psiquiátrica en México (Medina-Mora, Borges, Lara, Benjet, Blanco, Fleiz, Villatoro, Rojas, Zambrano, Casanova y Aguilar-Gaxiola, 2003), de este porcentaje las mujeres ocupan el 64.34%, lo que nos habla de la mayor frecuencia de síntomas depresivos entre las mujeres (Belló, Puentes-Rosas, Medina-Mora y Lozano, 2005), y la más alta proporción la relación del estrés con el suicidio en adolescentes (Gonzalez-Forteza, Villatoro, Pick y Collado, 1998; Hernández y Lucio, 2006), lo que concuerda con resultados de los estudios en otros países (Barra, Cerna, Kramm y Véliz, 2006).

Algunos estudios de adolescentes en México se han enfocado en la depresión, estrés, uso de drogas y tendencias suicidas. González Forteza, Salgado, Andrade y Tapia (1996), encontraron que las jóvenes adolescentes mexicanas del sur de la ciudad de México experimentan más estrés proveniente de los conflictos cotidianos que su contraparte, los varones, y ellas han de luchar con una mayor tendencia a la depresión.

Ante esta situación de riesgo de depresión en los adolescentes resulta importante identificar los niveles de depresión en esta población lo que nos lleva a la necesidad de desarrollar instrumentos que posean buenas cualidades psicométricas que evalúen la sintomatología depresiva entre los adolescentes a nivel poblacional.

La CES-D es una de las escalas para evaluar sintomatología depresiva más utilizadas en los ámbitos clínico y de investigación (Barra, Cerna, Kramm y Veliz, 2006; Calvete y Cardeñoso, 1999; Haringsma, Engels, Beekman, & Spinhoven, 2004; Mc Cauley, Pedroza, Brown, Boake, Levin, Goodman, & Merritt, 2006), fue desarrollada para evaluar sintomatología depresiva en las poblaciones clínica y general, estableciendo una clara distinción entre los pacientes depresivos y la población en general (Radloff, 1977). En México la CES-D ha sido adaptada a nuestra población y utilizada para estudios epidemiológicos en diversos grupos y con adolescentes en particular (González-Forteza, 1996; Mariño, Medina-Mora, Chaparro y González-Forteza, 1993; Salgado y Maldonado, 1994; Vallejo, Osorno y Mazadiego, 2008b)

Entre los estudios acerca de la depresión en adolescentes mexicanos utilizando la escala CES-D, podemos considerar los realizados en la Ciudad de México por González Forteza, 1996, González Forteza, Salgado, Andrade Palos, y Tapia, 1996; en Cuauhtitlan, Itzcalli, Estado de México y Saltillo, Coahuila por Díaz-Guerrero, 1984; con adolescentes del norte de Veracruz por Vallejo, Rodríguez, Osorno, Mazadiego, y Soto, 2006; Vallejo, Osorno y Mazadiego, 2008a, 2008b; y con adolescentes de origen mexicano residentes en Estados Unidos, por Katragadda y Tidwell (1998).

Las técnicas de funcionamiento diferencial de los ítems (DIF) permiten evaluar la validez de los instrumentos de medida. El análisis del DIF o del posible sesgo de los ítems de un test, hoy es parte del estudio de la validez de los instrumentos de medida (Zenisky, Hambleton & Robin, 2003). Se ha producido un acelerado desarrollo de teorías y métodos de detección debido a sus repercusiones sociales, políticas, educativas y psicológicas (Clauser, 2000; Fidalgo, 1996a, Gómez e Hidalgo, 1997; Millsap & Everson, 1993 y Robie, Muller & Champion, 2001) El test presenta sesgo (o DIF), si los puntajes de algunos de sus ítems no dependen sólo del nivel del sujeto en el constructo medido, sino que están influenciados por el grupo que provienen, entonces estos ítems no son válidos para medir a sujetos que pertenecen a grupos diferentes (Hambleton, Clauser, Mazor & Jones, 1993; Sireci, Harter, Yang, & Bhola, 2003).

Los métodos de detección del DIF, el Estadístico de Mantel-Haenszel y el de Regresión Logística son ampliamente utilizados. El estadístico de Mantel-Haenszel (MH) fue traído al análisis del DIF por Holland y Thayer (1988), trata de comparar la ejecución de un ítem entre dos grupos a través de distintos niveles en un determinado criterio, asumiendo que en cada nivel los sujetos de uno y otro grupo son comparables, por lo que si no hay DIF ambos lo ejecutarán por igual. Su lógica consiste en que si el ítem no presenta DIF, la proporción entre el número de personas que aciertan al ítem y las que lo fallan, será la misma en los dos grupos comparados a lo largo de todos los niveles de puntuación (Fidalgo, 1996b; Gómez e Hidalgo, 1997). Mazor, Clauser y Hambleton (1994) propusieron una modificación al procedimiento que consiste en calcular separadamente el estadístico para la población con puntuaciones más bajas y con las puntuaciones mayores en el test, lo que ha resultado una solución eficaz y viable en la detección del DIF no uniforme.

El procedimiento de Regresión Logística (RL) Fue propuesto por Swaminathan y Rogers (1990), es en un procedimiento de regresión donde la variable dependiente o criterio es de naturaleza categórica donde el nivel de habilidad del sujeto, la pertenencia a Grupo Focal (o de interés, GF) o Grupo de Referencia (grupo conocido, GR) y su interacción, son las variables independientes, que pueden tener naturaleza categórica o continua; cuando la VI es también categórica la relación entre VD y VI es logística. Para la estimación de la habilidad se utiliza la puntuación total del test. Mediante esta técnica se trata de determinar si en la función matemática necesaria para predecir las respuestas dicotómicas a un ítem se debe incluir un término referido a la interacción entre el grupo y la habilidad (DIF no uniforme), o al grupo de pertenencia (DIF uniforme), o bien si puede predecirse el nivel de habilidad de los sujetos con independencia de su grupo de pertenencia (ausencia de DIF). Gómez y Navas (1996) propusieron un sistema de purificación paso a paso de la habilidad lo que mejoró significativamente su poder de detección del DIF evitando falsos positivos.

La CES-D ha mostrado la presencia de DIF de género en otras culturas (Gelin y Zumbo, 2003; Zumbo, Gelin, & Hublely, 2001) por lo que es importante

evaluar la presencia de DIF de género entre los mexicanos en la CES-D por lo que el objetivo de este trabajo fue evaluar la existencia de DIF en cuanto a género de la versión en español de la CES-D, analizada mediante las técnicas de Mantel-Haenszel y de Regresión Logística con una muestra de adolescentes mexicanos.

## **Método**

### *Participantes*

La muestra se conformó por 1011 adolescentes en edades de 11 a 19 años, 486 mujeres y 525 hombres, 515 de la etnia totonaca y 496 urbanos estudiantes de seis escuelas secundarias y preparatorias públicas de la ciudad de Poza Rica y de las comunidades de La Guásima y Polutla del municipio de Papantla en el Estado de Veracruz. El nivel socioeconómico de dicha muestra fue bajo y medio. Todos alumnos regulares en sus centros educativos.

### *Instrumento*

Se utilizó la CES-D (Radloff, 1977) en su versión en español adaptada por Gonzalez-Forteza y un grupo de investigadores del Instituto Nacional de Psiquiatría (Mariño, Medina-Mora, Chaparro y González-Forteza, 1993). Es una escala de autorreporte compuesta de 20 ítems, que evalúan afecto depresivo, afecto positivo, síntomas somáticos y relaciones interpersonales; se responde en una escala tipo Likert de cuatro opciones (0, 1 a 2 días, 3 a 4 y de 5 a 7 días) donde la persona señala la ocurrencia de cada síntoma durante la semana pasada, su valor teórico fluctúa entre 0 y 60.

### *Procedimiento*

#### *Aplicación del test*

Se aplicó la escala siguiendo las instrucciones de Radloff (1977) con la ayuda de estudiantes de los últimos semestres de psicología debidamente entrenados

para ello. Las aplicaciones fueron hechas durante los horarios de clases de los estudiantes, mismos que se acordaron con las autoridades del plantel. Se solicitó la ayuda voluntaria y se garantizó a los estudiantes la confidencialidad de sus respuestas, se aclararon las dudas que surgieron durante la aplicación.

### Análisis del DIF

Para este análisis se consideró como GR a las mujeres, dado que en las investigaciones éstas tienden a obtener puntuaciones más altas que los varones en depresión y GF a los varones. Los dos métodos utilizados utilizan métodos de purificación de la variable de contraste. Para aplicar las técnicas se dicotomizaron las variables (López-Pina, 2005).

Para el estadístico de Mantel-Haenszel se utilizó el programa MHDIF elaborado por Fidalgo (1994) en la Universidad de Oviedo, utilizando la opción de DIF no uniforme. Se corrió el programa con el Método de división por mitades de Mazor, Clauser y Hambleton (1994), con el que se mejora la identificación de los ítems con DIF no uniforme respecto del método de Holland y Thayer (1988).

Para el cálculo de la Regresión Logística se utilizó el programa SPSS 15.0 con su opción *Regresión logística* utilizando el procedimiento Swaminathan y Rogers (1990) con el método de purificación de Gómez y Navas (1996).

## Resultados

### *Detección del DIF con Mantel-Haenszel (MH)*

Se identificaron los siguientes ítems:

Con DIF en baja ejecución: 9, 10, 17 y 19 (20%)

Con DIF en alta ejecución: 3, 6, 10, 14, 17 y 18 (30%)

En las Tablas 1 y 2 se relacionan los ítems con DIF con sus puntuaciones de Alfa y Chi cuadrada y Delta para las ejecuciones baja (1), y alta (2) en total son 8, es decir el 40% del total del test. Como se puede observar mediante los puntajes de  $\alpha$  todos mayores a 1, los ítems favorecen al GR, es decir, a las mujeres.

**Tabla 1. Detección de ítems con DIF en MH en baja ejecución**

Ítem	Alfa	Chi cuadrada	Delta
3. Sentía que no podía quitarme la tristeza, ni con la ayuda de mi familia o amigos(as)	1.43	1.93 no	-0.84
6.Me sentí deprimido(a)	1.55	2.76 no	-1.02
9. Pensé que mi vida era un fracaso	2.05	5.27 si	-1.69
10. Tenía miedo	2.40	8.31 si	-2.06
14. Me sentía solo(a)	1.39	1.87 no	-0.77
17.Lloraba a ratos	3.77	14.46 si	-3.12
18. Me sentía triste	1.26	0.71 no	-0.55
19. Sentía que no le caía bien a los demás	1.78	5.67 si	-1.36

**Tabla 2. Detección de ítems con DIF en MH en alta ejecución**

Ítem	Alfa	Chi cuadrada	Delta
3. Sentía que no podía quitarme la tristeza, ni con la ayuda de mi familia o amigos(as)	2.17	5.51 si	-1.82
6.Me sentí deprimido(a)	2.43	9.39 si	-2.08
9. Pensé que mi vida era un fracaso	1.20	0.35 no	-0.44
10. Tenía miedo	3.01	12.91 si	-2.59
14. Me sentía solo(a)	1.80	4.80 si	-1.39
17.Lloraba a ratos	2.31	7.47 si	-1.97
18. Me sentía triste	1.88	4.02 si	-1.48
19. Sentía que no le caía bien a los demás	1.05	0.01 no	-0.12

En la Tabla 3 se relacionan los ítems detectados con DIF en baja y alta ejecución e identificando si el DIF es de tipo uniforme o no uniforme.

**Tabla 3.Ítems detectados con DIF con MH y el tipo de DIF**

Ítem	Ej. Baja	Ej. Alta	DIF
3. Sentía que no podía quitarme la tristeza, ni con la ayuda de mi familia o amigos(as)		X	NOU*
6.Me sentí deprimido(a)		X	NOU
9. Pensé que mi vida era un fracaso	X		NOU
10. Tenía miedo	X	X	UNIF
14. Me sentía solo(a)		X	NOU
17.Lloraba a ratos	X	X	UNIF**
18. Me sentía triste		X	NOU
19. Sentía que no le caía bien a los demás	X		NOU

\*DIF no uniforme, \*\*DIF uniforme.

Como se puede apreciar la mayoría de los ítems poseen DIF no uniforme y solo el ítem 17 presenta DIF uniforme.

#### *Detección del DIF mediante el Análisis de la Regresión Logística (RL)*

Se identificaron 12 ítems con DIF (60%) de los 20 que contiene la prueba. La Tabla 4 presenta los ítems detectados con DIF con sus valores de  $\tau$  que muestran los ítems con DIF de mayor a menor intensidad, el  $\tau_2$  indica DIF uniforme, es decir, que la variable género se distribuye uniformemente en los valores altos y bajos de

la sintomatología depresiva y  $\tau_3$  indica DIF no uniforme, es decir, la interacción de las variables de género y sintomatología depresiva.

**Tabla 4. Ítems de la CES-D detectados con RL con DIF de género uniforme y no uniforme.**

Ítem	$\tau_2$		$\tau_3$		DIF
	Wald	Signif.	Wald	Signif.	
17. Lloraba a ratos	156.80	.000	17.08	.000	NOU*
6. Me sentí deprimido(a)	150.20	.000	11.67	.001	NOU
3. Sentía que no podía quitarme la tristeza, ni con la ayuda de mi familia o amigos(as)	145.70	.000	7.35	.007	NOU
10. Tenía miedo	147.37	.000	21.13	.000	NOU
18. Me sentía triste	143.91	.000	12.22	.000	NOU
9. Pensé que mi vida era un fracaso	131.70	.000	3.88	.049	UNIF**
14. Me sentía solo(a)	122.38	.000	11.70	.001	NOU
19. Sentía que no le caía bien a los demás	118.20	.000	6.09	.014	NOU
11. Dormía sin descansar	98.02	.000	1.41	.235	UNIF
20. No podía "seguir adelante".	92.46	.000	1.40	.236	UNIF
2. No tenía hambre, ni tenía apetito	86.95	.000	7.80	.005	NOU
8. Veía el futuro con esperanza*	98.47	.000	.237	.626	UNIF

\*DIF no uniforme, \*\*DIF uniforme.

De los ítems 6 presentan DIF no uniforme lo que implica la interacción entre el género y la depresión y dos DIF no uniforme, la diferencia es uniforme en el nivel de depresión en todos los niveles. Los todos los ítems identificados con DIF favorecen al GR (mujeres), siendo sus puntuaciones en el test sistemáticamente más altas que la de los hombres.

#### *Concordancia entre los Métodos de MH y RL*

En la tabla 5 se comparan los ítems detectados por los métodos de RL y de MH y se obtiene el porcentaje de acuerdo entre ellos.

**Tabla 5 Concordancia de detección de DIF entre MH y RL.**

Ítem	RL	MH	ACUERDO
17. Lloraba a ratos	X	X	SI
6. Me sentí deprimido(a)	X	X	SI
3. Sentía que no podía quitarme la tristeza,	X	X	SI

ni con la ayuda de mi familia o amigos(as)			
10. Tenía miedo	X	X	SI
18. Me sentía triste	X	X	SI
9. Pensé que mi vida era un fracaso	X	X	SI
14. Me sentía solo(a)	X	X	SI
19. Sentía que no le caía bien a los demás	X	X	SI
11. Dormía sin descansar	X		NO
20. No podía "seguir adelante".	X		NO
2. No tenía hambre, ni tenía apetito	X		NO
8. Veía el futuro con esperanza*	X		NO

Tomando la fórmula clásica de confiabilidad de observaciones de acuerdos/ acuerdos + desacuerdos x 100= 66.66%. Lo que nos indica un muy buen porcentaje de acuerdos entre los métodos de detección. Los ítems en que acuerdan los métodos que presentan DIF son: 17, 6, 3, 10, 18, 9, 14 y 19 en total 8 ítems de los 20 que presenta la CES-D es decir el 40% de los ítems están sesgados y favorecen al GR (mujeres).

### Discusión

La CES-D es una de las escalas para evaluar la sintomatología depresiva más utilizadas en la investigación clínica y poblacional y ha mostrado buenas cualidades psicométricas de validez y confiabilidad (González-Forteza, 1996; Mariño, Medina-Mora, Chaparro y González-Forteza, 1993; Vallejo, Osorno y Mazadiego, 2008a). En las investigaciones con la CES-D las mujeres han puntuado consistentemente más altas (Belló, Puentes-Rosas, Medina-Mora y Lozano, 2005; Mariño, Medina-Mora, Chaparro y González-Forteza, 1993; Vallejo, Osorno y Mazadiego, 2008<sup>a</sup>, Vallejo, Osorno y Mazadiego, 2008b). Existen diferentes planteamientos que han tratado de dar cuenta de esto, aduciendo aspectos biológicos de las mujeres, la importancia que conceden a las relaciones interpersonales y a diferentes estilos de afrontamiento de los problemas, sin embargo, no se han estudiado a fondo las características de los instrumentos de medición de la sintomatología depresiva.

En esta investigación se propuso evaluar la validez de la CES-D mediante los métodos de detección del DIF Mantel-Haenszel y Regresión Logística, con los que se identificó que 8 de los 20 ítems de la escala presentan DIF de género de una

manera consistente (acuerdo entre métodos). Los resultados de esta investigación coinciden con los obtenidos por Gelin y Zumbo (2003) y Zumbo, Gelin, & Hubley (2001), quienes encontraron a través de las técnicas de DIF ítems que puntuaban sistemáticamente más altas a las mujeres en la CES-D.

Es de considerar que esta es la primera investigación publicada en México evaluando la validez de la CES-D mediante las técnicas de identificación del DIF con adolescentes, lo que nos permite identificar esos ítems sesgados mejorar el instrumento e indagar sobre las causas de su sesgo generando nuevas preguntas de investigación acerca de la medición de la depresión.

Los métodos de detección del DIF manifestaron un alto grado de concordancia, lo que nos habla de la firmeza en la identificación de los ítems detectados. El estadístico de MH identificó 8 ítems con DIF, el análisis de RL identificó 12 ítems con DIF y la totalidad de los ítems identificados con los dos métodos favorecen a las mujeres. En relación a estos resultados, se puede afirmar que el 40% de los ítems de la CES-D permiten sistemáticamente que las mujeres obtengan puntajes mayores que los hombres, lo que representa un problema de validez interna, por lo que es importante hacer una revisión de los ítems detectados con DIF, explorar las causas de su funcionamiento anómalo y realizar propuestas para mejorar la prueba.

El DIF detectado en la CES-D puede deberse a la infiltración de la cultura en la construcción de la escala como lo señala el estudio realizado por Iwata (2002), donde las diferencias de cultura medida a través de las técnicas de DIF fueron muy evidentes, los roles y expectativas que la cultura tiene para cada género, nos señala la multidimensionalidad de estos ítems con DIF. Es de tomar en cuenta que los puntajes altos de las mujeres aparecen utilizando diversos instrumentos como afirma el *Informe Mundial sobre la Salud de 2001* donde se señala la prevalencia de depresión en el mundo en los hombres de 1.9% y de 3.2 en las mujeres, en otros estudios se obtuvieron puntuaciones más altas en las mujeres, (González-Forteza, Salgado de Snyder, Andrade-Palos, y Tapia, 1996; Masten, Caldwell-Colbert, Williams, Jerome, Mosby, Barrios, y Helton, 2003) Se requiere estudiar a mayor detalle la forma cómo se manifiesta la depresión en los

adolescentes masculinos, para de esta manera incluir ítems en la escala que aborden aspectos importantes de cómo ellos manifiestan la depresión, en ese sentido el papel de los psicólogos clínicos es de suma importancia así como el uso de redes semánticas, ya que éstos pueden proporcionar información muy valiosa para el desarrollo de nuevos ítems. El mejoramiento de la escala conducirá a considerar si lo que se está detectando en la prueba son síntomas de depresión o estamos evaluando la expresión de características del rol femenino, es decir, el constructo que se pretende medir o una variable espuria.

En esta investigación se demostró que las técnicas de identificación del funcionamiento diferencial de los ítems son útiles para evaluar la validez de una prueba incluso donde otras técnicas para evaluar la validez como el análisis factorial exploratorio no permitieron identificar los problemas (Mariño, Medina-Mora, Chaparro y González-Forteza, 1993; Vallejo, Osorno y Mazadiego, 2008b). Lo que nos lleva a proponer que se utilicen las técnicas de detección del DIF de manera cotidiana en la evaluación de la validez de las pruebas utilizadas para cerciorarse de la validez interna del instrumento de medida.

## Referencias

- Allgood-Merton, B., Lewinsohn, P.M. y Hops, H. (1990). Sex differences and adolescent depression. *Journal of Abnormal Psychology, 99*, 55-63
- Barra, E., Cerna, R., Kramm, D., y Véliz, V. (2006). Problemas de salud, estrés, afrontamiento, depresión y apoyo social en adolescentes. *Terapia Psicológica, 24(1)*, 55-61.
- Belló, M., Puentes-Rosas, E., Medina-Mora, M.E., y Lozano, R. (2005). Prevalencia y diagnóstico de depresión en población adulta en México. *Salud Pública de México, 47(1)*, 4-11.
- Calvete, E., y Cardeñoso, O. (1999). Creencias y síntomas depresivos: Resultados preliminares en el desarrollo de una Escala de Creencias Irracionales abreviada. *Anales de Psicología, 15(2)*, 179-190.
- Díaz-Guerrero, R. (1984). Tristeza y psicopatología en México, *Salud Mental, 7(2)*, 3-9

- Estévez, E., Musitu, G., y Herrero, J. (2005). El rol de la comunicación familiar y del ajuste escolar en la salud mental del adolescente. *Salud Mental, 28(4)*, 81-89.
- Gelin, M., y Zumbo, B. (2003). Differential item functioning results may change depending on how an item is scored: An illustration with the center for epidemiologic studies depression scale. *Educational and Psychological Measurement, 63(1)*, 65-74.
- González-Forteza, C. (1996). *Factores protectores y de riesgo de depresión e intento de suicidio en adolescentes*. Tesis de doctorado en Psicología. México: Facultad de Psicología de la UNAM.
- González-Forteza, C., Salgado de Snyder, V.N., Andrade-Palos, P. y Tapia, A.J. (1996). Gender differences in daily stress and depressive symptomatology among Mexican early adolescents. *New Trends in Experimental & Clinical Psychiatry, 12 (1)*, 17-22
- Gonzalez-Forteza, C., Villatoro, J., Pick, S, y Collado, M. E. (1998). "Estrés psicosocial y su relación con las respuestas de enfrentamiento y el malestar emocional en una muestra representativa de adolescentes al sur de la ciudad de México: Análisis según su nivel socioeconómico. *Salud Mental, 2*, 37-45.
- Hambleton, R.K., Clauser, B.E., Mazor, K.M., & Jones, R.W. (1993). Advances in the detection of differentially functioning test items. *European Journal of Psychological Assessment, 9(1)*, 1-18.
- Haringsma, R., Engels, G., Beekman, A.T., & Spinhoven, P. (2004). The criterion validity of the Center for Epidemiological Studies Depression Scale (CES-D) in a sample of self-referred elders with depressive symptomatology. *International Journal of Geriatry and Psychiatry, 19(6)*, 558-563.
- Hernández, Q., Lucio, E. (2006). "Evaluación del riesgo suicida y estrés asociado en adolescentes estudiantes mexicanos." *Revista Mexicana de Psicología, 23*, 45-52.
- Iwata, N. (2002). Race/ethnicity and depressive symptoms in community-dwelling young adults: a differential item functioning analysis. *Psychiatry Research, 110(3)*, 281-289.
- Katragadda, C.P. y Tidwell, R. (1998). Rural Hispanic adolescents at risk for depressive symptoms. *Journal of Applied Social Psychology, 28 (20)*, 1916-1930.
- López-Pina, J.A. (2005). Ítems politómicos vs. dicotómicos: Un estudio metodológico. *Anales de Psicología (2)*, 21,339-344.

- Mariño, M., Medina-Mora, M., Chaparro, J., y González-Forteza, C. (1993). Confiabilidad y estructura factorial del CES-D en una muestra de adolescentes mexicanos. *Revista Mexicana de Psicología*, 19(2), 141-145.
- Masten, W., Caldwell-Colbert, A., Williams, V., Jerome, W., Mosby, L. Barrios, Y. y Helton, J. (2003). Gender differences in depressive symptoms among Mexican adolescents. *Anales de Psicología*. (19) 1, 91-95
- Mazor, K.M., Clauser, B.E., & Hambleton, R.K. (1994). Identification of nonuniform differential item functioning using a variation of the Mantel-Haenszel procedure. *Educational and Psychological Measurement*, 54(2), 284-291.
- Mc Cauley, S.R., Pedroza, C., Brown, S.A., Boake, C., Levin, H.S., Goodman, H.S., & Merritt, S.G. (2006). Confirmatory factor structure of the Center for Epidemiologic Studies-Depression scale (CES-D) in mild-to-moderate traumatic brain injury. *Brain Injury*, 20 (5), 519-527.
- Medina-Mora, M.E., Borges, G., Lara, C., Benjet, C., Blanco, J., Fleiz, C., Villatoro, J., Rojas, E., Zambrano, J., Casanova, L., y Aguilar-Gaxiola, S. (2003). Prevalencia de trastornos mentales y uso de servicios: Resultados de la Encuesta Nacional de Epidemiología Psiquiátrica en México. *Salud Mental*, 26(4), 1-16.
- Radloff, L.S. (1977). The CES-D Scale: A Self-Report Depression Scales for Research in the General Population. *Applied Psychological Measurement*, (1), 3, 385-401.
- Salgado, N., y Maldonado, M. (1994). Características psicométricas de la Escala de Depresión del Centro de Estudios Epidemiológicos en mujeres mexicanas adultas de áreas rurales. *Salud Pública*, 36(2), 200-209.
- Swaminathan, H. & Rogers, H. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement*, 27, 361-370.
- Vallejo, A., Rodríguez, C., Osorno, R., Mazadiego, T., y Soto, P. (2006). Niveles de depresión en una muestra de adolescentes veracruzanos. *Revista Mexicana de Psicología, Suplemento especial*, pág. 234.
- Vallejo, A., Osorno, R., y Mazadiego, T. (2008a). Estilos parentales y sintomatología depresiva en una muestra de adolescentes veracruzanos. *Enseñanza e Investigación en Psicología*, 13(1) ,91-105.
- Vallejo, A., Osorno, R., y Mazadiego, T. (2008b). Sintomatología depresiva medida con la CES-D en adolescentes del norte de Veracruz. En *Investigaciones*

*Psicosociales del Adolescente 2007*. Tuxtla Gutiérrez: Escuela de Psicología de la Universidad de Ciencias y Artes de Chiapas, (en prensa).

Washburn, J.; Hillman, S. & Sawiloswsky, S. (2004). Gender and gender- role orientation differences on adolescent's coping with peer stressors. *Journal of Youth and Adolescence*. XXXIII, 1, 31-40.

Zenisky, A., Hambleton, R.K., & Robin, F. (2003). Detection of differential item functioning in large-scale state assessments: A study evaluating a two-stage approach. *Educational and Psychological Measurement*, 63(1), 51-64.

Zumbo, B., Gelin, M. & Hubley, A. (2001). *Psychometric Study of the CES-D: Factor Analysis and DIF*. Conferencia presentada en la International Neuropsychological Society's 29th Annual Meeting, University of British Columbia presentada el 16 de Febrero de 2001. Recuperado el 20 de Febrero de 2008 de <http://educ.ubc.ca/faculty/zumbo/ins2001/INS2001.pdf>.