



Revista Mexicana de Trastornos Alimentarios

Mexican Journal of Eating Disorders

Journal homepage: <http://journals.iztacala.unam.mx/>

Eating disorders questionnaires validated in Mexican women: A review of the literature

Instrumentos para trastornos del comportamiento alimentario validados en mujeres mexicanas: Una revisión de la literatura

Karina Franco Paredes¹, Georgina Leticia Alvarez Rayón², Rosa Elena Ramírez Ruelas¹

¹ Centro de Investigaciones en Comportamiento Alimentario y Nutrición (CICAN), CUSur-Universidad de Guadalajara, Jalisco, México.

² Laboratorio de trastornos del comportamiento alimentario, FES-Iztacala, Universidad Nacional Autónoma de México, México.

Abstract

The purpose of this review was to analyze the studies about psychometric properties of eating disorders (ED) questionnaires validated in Mexican women. We search papers published in scientific journal about psychometric properties of questionnaires to measure symptoms or characteristics associated with ED. Seventeen papers that evaluated the reliability or validity of 16 questionnaires were identified (11 were created in another countries and five were created by Mexican researchers). It was found that questionnaires have adequate reliability; however, the researchers only analyzed the internal consistency. Several methods were employed to analyze the validity of the questionnaires; the most common were factorial analysis exploratory and homogeneity of the scale. There are seven questionnaires with cut-off point, which permit to identify people at risk for ED. It was concluded that it is necessary that future researches provide further evidences about psychometric properties of the revised questionnaires.

Resumen

El propósito de la presente revisión fue hacer un análisis de aquellos estudios en que se han examinado las propiedades psicométricas de instrumentos de evaluación de trastornos del comportamiento alimentario (TCA) en mujeres mexicanas. Se buscaron artículos publicados en revistas científicas sobre propiedades psicométricas de instrumentos para medir síntomas o características asociadas a los TCA. Se localizaron 17 publicaciones en las que se evaluó la confiabilidad y/o validez de 16 cuestionarios de autoinforme (11 creados en otros países y cinco construidos por investigadores mexicanos). Se encontró que los instrumentos poseen confiabilidad adecuada, sin embargo, sólo se ha analizado la consistencia interna. Para evaluar la validez de los instrumentos se han empleado diferentes procedimientos, siendo el más común el análisis factorial exploratorio y el análisis de la homogeneidad de la prueba. Siete de los instrumentos cuentan con punto de corte, el cual permite identificar a personas en riesgo de desarrollar algún TCA. Se concluyó que es necesario realizar más investigaciones que fortalezcan la evidencia existente sobre las propiedades psicométricas de los instrumentos revisados.

INFORMACIÓN ARTÍCULO

Recibido: 13/10/2011
Revisado: 09/12/2011
Aceptado: 12/12/2011

Key words: Assessment, questionnaires, eating disorders, women, Mexico.

Palabras clave: Evaluación, instrumentos, trastornos del comportamiento alimentario, mujeres, México.

Introducción

Los trastornos de la conducta alimentaria (TCA) son procesos psicopatológicos que suponen severas modificaciones en las actitudes y los comportamientos relacionados con la ingestión de alimento, como resultado de una distorsión en la percepción de las dimensiones corporales y un miedo intenso a la obesidad (Mancilla-Díaz et al., 2006). De acuerdo con el Manual Diagnóstico y Estadístico de los Trastornos Mentales, cuarta edición, texto revisado (DSM-IV-TR, por sus siglas en inglés; American Psychiatry Association [APA], 2000) existen dos TCA específicos: la anorexia nerviosa (AN) y la bulimia nerviosa

Correspondencia: Karina Franco Paredes, CUSur-Universidad de Guadalajara. Av. Enrique Arreola Silva 883, C.P. 49000, Ciudad Guzmán, Jalisco, México. Correo electrónico: karina.franco@cusur.udg.mx

(BN), y otros denominados trastornos de la conducta alimentaria no especificados (TCANE), en los cuales se incluye el trastorno por atracón (TPA).

En México se realiza investigación sobre TCA desde hace aproximadamente dos décadas, el conocimiento generado en este periodo fue presentado de manera organizada en dos trabajos de revisión (Mancilla et al., 2006; Unikel & Bojorquez, 2007); ambos documentos incluyen una descripción de los avances en diferentes líneas de investigación, entre las que destaca la referente a la evaluación de los TCA. Dentro de las múltiples herramientas que pueden emplearse con este fin, los instrumentos de autoinforme representan una opción que ha proporcionado resultados satisfactorios, por tanto, su uso se ha

generalizado tanto en el campo de la investigación como en la práctica clínica.

Una de las primeras necesidades que debieron solventar los investigadores pioneros en el estudio de los TCA en México, fue contar con instrumentos de medición que demostraran su aplicabilidad en muestras mexicanas. Ante esta demanda, se han seguido dos caminos: el primero implica la traducción y/o adaptación de instrumentos de amplio uso en el mundo; mientras que el segundo camino comprende la construcción de instrumentos que, aunque basados –en mayor o menor medida– en alguno ya existente, pretenden atender a las características idiosincráticas de nuestro país. Resulta innegable que la línea de investigación sobre evaluación ha constituido un pilar para la producción de conocimiento en cuanto a los TCA en México. Por tanto, el propósito de la presente revisión fue analizar aquellos estudios en que se han examinado las propiedades psicométricas de instrumentos de evaluación de TCA en mujeres mexicanas.

Método

La búsqueda inicial se basó en el listado de referencias de dos revisiones previas de la investigación realizada en México sobre TCA (Mancilla et al., 2006; Unikel & Bojorquez, 2007). Se identificaron aquellos artículos que versaban sobre el análisis de propiedades psicométricas de instrumentos de evaluación y los respectivos artículos de la validación original. Posteriormente se procedió a localizar los artículos identificados, así como otros que no estaban incluidos en las revisiones antes citadas, en dos bases de datos especializadas (*Redalyc* y *EBSCO*), así como en buscadores de internet (e.g. google académico). Finalmente, con base a la técnica *bola de nieve*, también se ubicaron aquellos artículos referenciados en las publicaciones obtenidas en la búsqueda previa. Cabe señalar que esta revisión sólo incluyó estudios publicados en revistas especializadas.

Resultados

Se localizaron 17 publicaciones que fueron clasifica-

das considerando dos criterios: 1. Si correspondían a instrumentos traducidos y/o adaptados para México, o bien, si fueron contruidos para su empleo en población mexicana; y 2. Objetivo del instrumento, identificando tres distintos: tamizar o cribar, medir síntomas o características específicas de TCA, o factores de riesgo. Por tanto, en la presente revisión la información se organizó considerando ambos criterios.

Instrumentos traducidos y/o adaptados

En este apartado se describen los estudios que han examinado las propiedades psicométricas de 11 instrumentos que fueron contruidos en otros países, por lo que en todos los casos se inicia exponiendo los datos correspondientes a la fuente original.

Instrumentos de tamizaje

Test de Actitudes Alimentarias (EAT-40, por sus siglas en inglés)

Objetivo: Medir síntomas y conductas características de la AN.

Descripción: Incluye 40 reactivos con seis opciones de respuesta en escala tipo Likert, que se califican de 0 a 3 puntos.

El EAT-40 fue creado y validado por Garner y Garfinkel (1979) en dos submuestras de mujeres diagnosticadas con AN ($n = 32$ y 33), de acuerdo a los criterios de Feighner et al. (1972), y en dos submuestras control ($n = 34$ y 59). La consistencia interna del EAT-40 fue $Alpha = .94$ en la muestra total, y $.79$ en las mujeres con AN. La puntuación total del instrumento diferenció entre las mujeres con AN y las mujeres control en la dirección esperada; además, al analizar la predicción diagnóstica se estableció que 30 era un punto de corte adecuado para fines de tamizaje. En un segundo trabajo, Garner, Olmstead, Bohr y Garfinkel (1982) analizaron la estructura factorial del EAT-40 con una muestra de 160 mujeres con AN, derivando tres factores que agruparon 26 ítems y explicaron 40.2% de la varianza total: Dieta, Bulimia y Preocupación por la Comida, y Control Oral. Factores que mostraron consistencia interna adecuada en

la muestra clínica ($Alpha = .90, .84$ y $.83$, respectivamente), pero no así en mujeres control ($Alpha = .86, .61$ y $.46$, respectivamente), dado que en los dos últimos factores fue deficiente. Esta estructura factorial dio origen a la versión corta del EAT-40, denominada EAT-26, misma que fue adaptada para su empleo en niños, resultando la Children's version of EAT (ChEAT; Maloney, McGuire & Daniels, 1988), la cual fue validada en 318 niños y niñas. La consistencia interna fue $Alpha = .76$ en la muestra total, y por grado escolar -de tercero a sexto- fluctuó entre $.68$ y $.80$; en tanto que la confiabilidad test-retest fue $r = .81$ en una submuestra de 68 niños. En general, tanto el EAT-40 como el ChEAT han mostrado poseer propiedades psicométricas adecuadas para medir las actitudes y las conductas características de la AN.

En México se ha examinado la confiabilidad y validez tanto del EAT-40 como del ChEAT. La versión en español del EAT-40 (Castro, Toro, Salamero & Guimerá, 1991) fue adaptada por Alvarez-Rayón et al. (2004) y aplicada a una muestra clínica de 276 mujeres diagnosticadas con TCA (AN = 52, BN = 102 y TCANE = 122), de acuerdo a los criterios diagnósticos del DSM-IV (APA, 1994), y a una muestra control de 280 mujeres. La consistencia interna del instrumento fue $Alpha = .90$ en la muestra con TCA y $.93$ en la muestra total, coeficientes similares a los reportados por Garner y Garfinkel (1979). El análisis factorial exploratorio (AFE) derivó cinco factores que agruparon 25 reactivos y explicaron 46.6% de la varianza: Restricción Alimentaria ($Alpha = .88$), Bulimia ($Alpha = .87$), Motivación para Adelgazar ($Alpha = .85$), Preocupación por la Comida ($Alpha = .77$) y Presión Social Percibida ($Alpha = .76$). La puntuación total diferenció entre las mujeres con TCA y las mujeres control; además, la mejor predicción diagnóstica para AN o BN se obtuvo con el punto de corte de 28 (sensibilidad = 83% y especificidad = 91%), en tanto que para TCANE se propuso el punto de corte de 22 (83% y 82%, respectivamente). En general, se demostró que el EAT-40 cuenta con propiedades psicométricas adecuadas para su empleo en mujeres mexicanas; no obstante, es importante mencionar que se encontraron dos cambios

relevantes respecto al estudio original. Por un lado, la estructura factorial fue diferente, por otro lado, el punto de corte propuesto para México fue ligeramente menor al original.

El ChEAT fue validado en México por Escoto y Camacho (2008) en una muestra de 1241 púberes y adolescentes, de entre 9 y 15 años de edad. Se comprobó la homogeneidad del instrumento, debido a que todos los reactivos correlacionaron positiva y significativamente con la puntuación total; además, tanto la consistencia interna ($Alpha = .82$), como la confiabilidad test-retest ($r = .79$) fueron adecuadas. El análisis factorial derivó cinco factores que agruparon 25 reactivos y explicaron 43.74% de la varianza: Motivación para Adelgazar ($Alpha = .74$), Evitación de Alimentos Engordantes ($Alpha = .77$), Preocupación por la Comida ($Alpha = .67$), Conductas Compensatorias ($Alpha = .69$), y Presión Social Percibida ($Alpha = .63$). Cabe señalar que tres de estos factores coincidieron conceptualmente con los identificados en el EAT-40 por Alvarez et al. (2004), sin embargo, no agruparon los mismos reactivos. Finalmente, como evidencia de validez concurrente, se encontró que el ChEAT correlacionó significativamente con el Cuestionario de Imagen Corporal (BSQ, por sus siglas en inglés; Cooper, Taylor, Cooper & Fairburn, 1987). El estudio de Escoto y Camacho fundamenta que el ChEAT es un instrumento adecuado para medir actitudes y conductas alimentarias inadecuadas en púberes y adolescentes mexicanos. No obstante, su estructura factorial mostró variabilidad.

Test de Bulimia (BULIT, por sus siglas en inglés)

Objetivo: Evaluar y detectar en la población general a quienes presentan riesgo de desarrollar síntomas de BN.

Descripción: Tiene 36 reactivos con cinco opciones de respuesta que hacen referencia a intensidad o frecuencia, y se califican de 1 a 5. Para calcular la puntuación total se sugiere excluir cuatro ítems (7, 33, 34 y 36) que miden uso de laxantes, de diuréticos o amenorrea, debido a que son poco frecuentes entre la población general.

El BULIT fue desarrollado por Smith y Thelen

(1984) con base a un estudio de cuatro fases. En las dos primeras se construyó el instrumento y se propuso el punto de corte de 102, el cual clasificó correctamente entre mujeres con BN ($n = 20$) y control ($n = 92$), con valores superiores a 90% en especificidad y sensibilidad. El AFE derivó siete factores que explicaron 50% de la varianza: Conducta de Sobreingesta (actualmente denominada atracón, por lo que en lo sucesivo se empleará este término), Sentimientos Posteriores al Atracón, Conducta de Vómito, Preferencias Alimentarias durante el Atracón, Fluctuaciones en el Peso, Abuso de Laxantes, e Irregularidad en el Ciclo Menstrual. Esta estructura factorial fue considerada tentativa, debido al reducido tamaño de la muestra y porque cuatro de los siete factores contaban con dos ítems, por ende, propiamente no podrían considerarse como factores (Yela, 1997). En la tercera fase se encontró que, de acuerdo al punto de corte de 102, 4% de las mujeres presentaron riesgo de BN. En la cuarta fase se evaluó la confiabilidad test-retest a dos meses, y se encontró una correlación de $r = .87$ entre las puntuaciones. Los hallazgos de este estudio evidenciaron que el BULIT es un instrumento confiable y válido para identificar individuos en riesgo de BN entre la población general, ya que evalúa aspectos de la conducta de atracón, conductas purgativas, afecto negativo y fluctuaciones en el peso.

En México, Alvarez, Mancilla y Vázquez (2000) adaptaron y validaron la versión en español del BULIT (Mora, 1992). Para ello realizaron un estudio con dos fases, en la primera participaron 2168 mujeres y en la segunda 35 mujeres que cumplían criterios para BN o TCANE relacionado con BN, de acuerdo al DMS-IV (APA, 1994), y 35 mujeres control. La consistencia interna del BULIT para la muestra clínica fue $Alpha = .88$ en la primera fase y $.84$ en la segunda. Asimismo, se encontró que todos los reactivos del instrumento correlacionaron significativamente con la puntuación total. El AFE realizado en la primera fase derivó tres factores que agruparon 22 reactivos y explicaron 33.2% de la varianza total: Atracón ($Alpha = .85$), Sentimientos Negativos Posteriores al Atracón ($Alpha = .82$) y Conductas Compensatorias ($Alpha = .58$), resultando insatisfactoria la consis-

tencia interna de este último factor; además, tanto la puntuación total como los factores diferenciaron entre las mujeres con TCA y las control. Finalmente, en un estudio paralelo, Alvarez (2000) propuso el punto de corte de 85 (especificidad = 97% y sensibilidad = 79%). En conclusión, el BULIT mostró ser confiable y válido para la detección oportuna de personas en riesgo de presentar o desarrollar síntomas de BN.

Test de Investigación Bulímica de Edimburgo (BITE, por sus siglas en inglés)

Objetivo: Evaluar aspectos cognitivo-conductuales de la BN, con énfasis en la medición del atracón.

Descripción: Consta de 33 reactivos, 30 con un formato de respuesta dicotómico (No- Sí) que se califica con 0 y 1, mientras que los tres reactivos restantes tienen 5, 6 ó 7 opciones de respuesta en escala tipo Likert y se califican de 1 a 5, 6 ó 7, respectivamente.

El BITE fue elaborado por Henderson y Freeman (1987), quienes realizaron cuatro estudios para validarlo. En el primero encontraron que los reactivos del BITE diferenciaron entre mujeres con atracón ($n = 15$) y mujeres control ($n = 40$) en la dirección esperada. La puntuación de 25 se estableció como punto de corte para el total del instrumento, para la subescala Síntomas fue 20 y para la subescala Severidad fue 5. En el segundo estudio, la muestra de mujeres con BN ($n = 32$) obtuvo puntuaciones significativamente mayores que la muestra control ($n = 32$), tanto en el total del instrumento como en las dos subescalas; asimismo, se confirmó que las participantes con BN rebasaron los puntos de corte del instrumento. La consistencia interna de las subescalas, en la muestra total, fue: $Alpha = .96$ para Síntomas y $.62$ para Severidad. Como evidencia de validez concurrente se encontró que el BITE correlacionó significativamente con seis de las ocho subescalas del Inventario de Trastornos Alimentarios (EDI, por sus siglas en inglés; Garner, Olmstead & Polivy, 1983), excepto Perfeccionismo y Desconfianza Interpersonal, así como con tres factores del EAT-40. En el tercer estudio se encontró que el BITE fue sensible para detectar cambios producidos por el tratamiento en mujeres con BN ($n =$

27). Finalmente, en el cuarto estudio se encontró que la confiabilidad test-retest (a 15 semanas) fue $r = .86$ en mujeres con BN ($n = 10$) y mujeres control ($n = 30$). El estudio mostró que las propiedades psicométricas del BITE fueron adecuadas.

La adaptación y validación para México de la versión en español del BITE (Rivas, Bersabé & Jiménez, 2004) fue realizada por Franco et al. (2010), quienes trabajaron con muestras comunitarias provenientes de dos estados: Jalisco ($n = 172$) y Sonora ($n = 159$), y con una muestra de 73 mujeres diagnosticadas con TCA (BN = 23 y TPA = 50). Se encontró que 25 de los 33 reactivos obtuvieron correlaciones mayores a .30 con la puntuación total del instrumento en las tres muestras (Jalisco, Sonora y clínica); mientras que la consistencia interna fue $Alpha = .85$ en las mujeres de Jalisco, .90 en las de Sonora y .83 en la muestra clínica. La puntuación total del BITE diferenció entre la muestra clínica y comunitaria en la dirección esperada. Al analizar la predicción diagnóstica, se identificó que 21 fue el punto de corte conveniente para el total del instrumento, 5 para la subescala Síntomas y 6 para la subescala Severidad, con porcentajes adecuados de sensibilidad (84, 88 y 71, respectivamente) y especificidad (83, 86 y 63, respectivamente). Respecto a la validez concurrente, se encontró que el BITE correlacionó con la puntuación total del EAT-40 y cuatro de los factores derivados de la validación en México. Los hallazgos obtenidos mostraron que la consistencia interna del instrumento fue adecuada y que existe evidencia favorable respecto a su validez.

Cuestionario de Patrones de Alimentación y Peso-Revisado (QEWP-R, por sus siglas en inglés)

Objetivo: Identificar individuos con episodios de atracón recurrente, así como discriminar entre pacientes con TPA y BN.

Descripción: Contiene 13 reactivos con un formato de respuesta en escala dicotómica, que se califican de acuerdo a reglas de decisión propuestas por los autores.

La primera versión del QEWP fue creada por Spitzer et al. (1992), quienes encontraron que el instru-

mento diferenció entre personas que participaban en programas de control de peso, personas con atracón y comedores compulsivos, y que la consistencia interna fue $Alpha = .75$ para la muestra clínica y .79 para la muestra control. Posteriormente, Spitzer, Yanovsky y Marcus (1993) evaluaron las propiedades psicométricas de una versión revisada del cuestionario, a la que denominaron QEWP-R. Este instrumento se aplicó a una muestra de 1785 personas que asistían a programas de control de peso, una muestra control de 942 personas y una muestra clínica de 75 mujeres diagnosticadas con BN de acuerdo al DSM-IV (APA, 1994). La puntuación total del instrumento diferenció entre las tres muestras en la dirección esperada y se obtuvo un valor $Kappa = .60$ entre el diagnóstico realizado por especialistas y el resultado del instrumento. Los autores concluyeron que el QEWP-R resultó útil para identificar la frecuencia y duración del atracón, el uso de métodos para controlar el peso y el grado de ansiedad relacionada con el atracón.

El QEWP-R fue traducido, adaptado y validado en México por López, Mancilla, Vázquez, Alvarez y Franco (2011) en una muestra de 448 mujeres que asistían a centros de ejercicio aeróbico-reductivo, a instituciones educativas, o a solicitar evaluación por un posible TCA. El cuestionario clasificó correctamente a 83.3% de la muestra con TCA ($n = 39$) y a 72.5% de la muestra sin TCA. El resultado del QEWP-R correlacionó positivamente ($Kappa = .60$) con la Entrevista para el Diagnóstico de Trastornos Alimentarios (IDED, por sus siglas en inglés; Kutlesic, Williamson, Gleaves, Barbin & Murphy-Eberenz, 1998). Asimismo, se encontró que algunas preguntas del QEWP-R se asociaron con las puntuaciones en los tres factores del BULIT derivados para México. Estos hallazgos indican que el QEWP-R presenta adecuada validez de criterio y es recomendable como instrumento de tamizaje de TPA y BN.

Instrumentos que miden características específicas de los TCA

Escala de Atracón (BES, por sus siglas en inglés)

Objetivo: Evaluar características cognitivo-conduc-

tuales del atracón.

Descripción: Se compone de 16 reactivos, ocho miden aspectos conductuales del atracón y, los ocho restantes, los sentimientos y/o cogniciones asociadas al mismo. Cada reactivo incluye cuatro enunciados que miden la severidad del rasgo y se califican en una escala de 0 a 3 puntos.

El BES fue desarrollado por Gormally, Black, Daston y Rardin (1982), quienes aplicaron el instrumento en personas que acudían a tratamiento conductual para obesidad. Se encontró que la consistencia interna fue $Alpha .85$ y que la puntuación total discriminó entre personas con diferentes niveles de severidad del atracón, de acuerdo al juicio de expertos. Este estudio mostró que el instrumento tiene adecuada consistencia interna y resultó útil para diferenciar entre niveles de severidad del atracón. No obstante, existe poca evidencia respecto a su validez.

La validación del BES en población mexicana fue realizada por Zuñiga y Robles (2006) en una muestra de 81 mujeres y 19 varones. El AFE derivó dos factores que explicaron 56.25% de la varianza: Sentimientos y Cogniciones ($Alpha = .88$) y Manifestaciones Conductuales ($Alpha = .87$). Cabe señalar que si bien los autores no precisan el número de reactivos agrupados en los factores, de la inspección visual de la tabla correspondiente se deduce que fueron los 16 reactivos; no obstante, cuatro obtuvieron cargas factoriales altas en los dos factores y dos más se agruparon en un factor diferente al que originalmente pertenecían. Los hallazgos de este estudio sólo aportan evidencia respecto a la validez de constructo del instrumento, específicamente sobre la estructura factorial, por lo que es necesario realizar más investigaciones que demuestren la confiabilidad y validez de criterio del BES.

Cuestionario de Imagen Corporal (BSQ, por sus siglas en inglés)

Objetivo: Evaluar la presencia de preocupación e insatisfacción con la imagen corporal.

Descripción: Incluye 34 reactivos con seis opcio-

nes de respuesta en escala tipo Likert, que se califican de 1 a 6.

El BSQ fue creado y validado por Cooper et al. (1987) con mujeres diagnosticadas con BN ($n = 38$), de acuerdo a los criterios del DSM-III-R (APA, 1987), y tres muestras control ($n = 331, 119$ y 85). La puntuación total del cuestionario diferenció entre las muestras en la dirección esperada: muestra clínica vs. control, mujeres preocupadas por su imagen corporal vs. mujeres que no estaban preocupadas, y probables casos de BN vs. muestra control. Como evidencia de validez concurrente, se informó que el BSQ correlacionó con la subescala insatisfacción corporal del EDI y con la puntuación total del EAT-40. Los autores concluyen que el BSQ es un cuestionario que proporciona información importante para la detección de posibles casos de TCA.

La versión en español del BSQ (Raich et al., 1996) fue adaptada y validada en México por Vázquez-Arévalo et al. (2011) en una muestra de 216 mujeres diagnosticadas con TCA (AN-R restrictiva = 22, AN-Purgativa = 17, BN-Purgativa = 87 y TCANE = 90), de acuerdo al DSM-IV-TR (APA, 2000), y una muestra control de 256 mujeres. La puntuación total y de cada reactivo fue significativamente mayor en la muestra clínica que en la muestra control. La consistencia interna del instrumento fue $Alpha = .98$ para la muestra total; además, todos los reactivos correlacionaron positiva y significativamente con la puntuación total. El AFE derivó dos factores que agruparon 20 reactivos y explicaron 63.8% de la varianza total: Malestar Corporal Normativo ($Alpha = .95$) y Malestar Corporal Patológico ($Alpha = .94$). Asimismo, se estableció el punto de corte de 110, con sensibilidad y especificidad de 84%. También se confirmó que el BSQ discriminó y clasificó a mujeres que presentan algún TCA de quienes no lo presentan, discriminando mejor entre TCA específicos y muestra control. Considerando los hallazgos descritos, puede decirse que el BSQ es un instrumento altamente confiable, ya que mide de manera precisa la insatisfacción corporal, la cual es uno de los principales factores de riesgo proximales de los TCA.

Instrumentos que miden factores asociados a los TCA

Inventario de Trastornos Alimentarios (EDI, por sus siglas en inglés)

Objetivo: Evaluar características cognitivo-conductuales asociadas a la AN y la BN.

Descripción: Está conformado por 64 reactivos con seis opciones de respuesta en escala tipo Likert, que se califican de 0 a 3.

El EDI fue construido y validado por Garner et al. (1983) en una muestra de 113 mujeres diagnosticadas con AN, de acuerdo a los criterios de Feighner et al. (1972), y 557 mujeres universitarias (muestra control). Los 64 reactivos diferenciaron significativamente entre las muestras, en tanto que 61 tuvieron una correlación mayor a .40 con la puntuación total. Los autores propusieron que el EDI evaluaba ocho características (factores): Motivación para Adelgazar, Bulimia, Insatisfacción Corporal, Ineficacia, Perfeccionismo, Desconfianza Interpersonal, Identificación Interoceptiva y Miedo a Madurar, cuya consistencia interna fluctuó entre $Alpha = .82$ (Perfeccionismo) y $.90$ (Bulimia, Insatisfacción Corporal e Ineficacia) en la muestra clínica, en tanto que en la muestra control fue desde $.65$ (Miedo a Madurar) hasta $.91$ (Insatisfacción Corporal). Los factores del EDI clasificaron correctamente a 85% de la muestra total. También se encontró que la puntuación total diferenció a mujeres con AN vs. con BN ($n = 195$), mujeres con obesidad ($n = 44$) con historia de obesidad ($n = 52$), mujeres control vs. varones control ($n = 166$), y mujeres con AN vs. mujeres recuperadas de AN. En cuanto a la validez concurrente, se observó que los factores del EDI correlacionaron con diversas medidas de aspectos psicológicos (e.g. AN, restricción alimentaria, insatisfacción corporal, locus de control, autoestima, depresión y sensibilidad interpersonal). Los autores concluyeron que el inventario presenta adecuada consistencia interna y validez. Posteriormente se llevó a cabo una revisión del EDI y se adicionaron 27 reactivos para evaluar tres aspectos más que suelen asociarse a los TCA (ascetismo,

impulsividad e inseguridad social), y fue así como se generó una nueva versión del instrumento, el EDI-2 (Garner, 1991).

Se tiene conocimiento de tres investigaciones respecto al análisis de la confiabilidad y validez del EDI en muestras mexicanas, dos sobre la versión original del EDI y una sobre el EDI-2. En primera instancia, Mancilla, Franco, Alvarez y Vázquez (2003) adaptaron y validaron la versión en español del EDI (Guimerá & Torrubia, 1987), y la aplicaron en tres muestras de mujeres (comunitaria = 669, TCA = 99 y control = 90). La consistencia interna fue $Alpha = .87$, $.92$ y $.78$, respectivamente; además, todos los reactivos del EDI, excepto el 1, correlacionaron con la puntuación total. Por medio del AFE se identificaron ocho factores que agruparon 44 reactivos y explicaron 41.7% de la varianza total; estructura que coincidió con la propuesta por Garner et al. (1983). En la muestra comunitaria, la consistencia interna de los factores fue desde $Alpha = .47$ (Perfeccionismo) hasta $.83$ (Motivación para Adelgazar); en la muestra clínica, desde $.40$ (Perfeccionismo) hasta $.86$ (Bulimia). La puntuación total del EDI y sus ocho factores, excepto Perfeccionismo, diferenciaron entre las mujeres con TCA y las control; asimismo, la puntuación total clasificó correctamente a 81.5% de las mujeres de la muestra total. Finalmente, se encontró que esa puntuación y la de seis de sus factores correlacionaron con el total del EAT y algunos de sus factores derivados para México. Si bien los autores concluyeron que la puntuación total del EDI y la mayoría de sus factores presentaron adecuada consistencia interna, puntualizan que la correspondiente a Perfeccionismo fue inaceptable en las tres muestras, además de que no diferenció entre las mujeres con TCA y las control.

Un segundo trabajo de investigación sobre el EDI fue realizado por Unikel, Bojorquez, Carreño y Caballero (2006), quienes trabajaron con una muestra de 523 mujeres con diagnóstico de TCA (5.7% AN-Purgativa, 8.0% AN-Restrictiva, 45.1% BN-Purgativa y 41.3% TCANE), de acuerdo al DSM-IV (APA, 1994). Evaluaron la homogeneidad del instrumento y encontraron que 52 de los 64 reactivos del EDI co-

rrelacionaron con la puntuación total. El AFE derivó seis factores que agruparon 36 reactivos y explicaron 56% de la varianza total: Bulimia, Motivación para Adelgazar, Conciencia interoceptiva, Ineficacia, Miedo a Madurar e Insatisfacción Corporal. La consistencia interna del EDI, considerando los 36 reactivos agrupados fue $Alpha = .92$, en tanto que para los factores fluctuó entre .76 (Miedo a Madurar) y .86 (Bulimia). Como evidencia de validez concurrente, el instrumento correlacionó positivamente con la Symptom Check List (Derogatis, Lipman & Covi, 1973) y negativamente con el Inventario de Autoestima de Coopersmith (Coopersmith, 1967). Los autores concluyeron que el EDI, en la versión de 36 reactivos, es adecuado para evaluar las características psicológicas de personas mexicanas con TCA, excepto por los factores Perfeccionismo y Desconfianza Interpersonal. En general, las dos investigaciones en las que se analizaron las propiedades psicométricas del EDI aportaron evidencias a favor de su confiabilidad y validez. No obstante, la estructura factorial mostró inconsistencias debido a que se agrupa un número diferente de reactivos y no se derivan los mismos factores. Asimismo, el factor Perfeccionismo resultó inadecuado en ambos estudios. Por tanto, es necesario realizar investigaciones que contribuyan a definir la estructura factorial del EDI.

En lo que refiere al EDI-2, éste fue traducido y validado en mujeres mexicanas por García-García, Vázquez-Velázquez, López-Alvarenga y Arcila-Martínez (2003). El instrumento fue aplicado en 22 mujeres diagnosticadas con TCA, de acuerdo a los criterios del DSM-IV (APA, 1994), y 25 mujeres sin TCA. La consistencia interna fue $Alpha = .98$ para el total de la muestra, .94 tanto para las mujeres con TCA como para las mujeres control, y mayor a .85 en los 11 factores propuestos por Garner (1991); asimismo, se encontró que la puntuación total y de cada factor diferenció entre las muestras. Los autores seleccionaron dos puntos de corte para el EDI-2 y cada factor, uno que representó mayor sensibilidad y otro con mayor especificidad. En aquellos con mayor sensibilidad, ésta fluctuó entre 68 y 91%, con especificidad de 68-84%; en tanto que en los puntos de corte

con mayor especificidad, ésta fluctuó entre 76 y 92%, con especificidad de 68-86%. Los autores concluyeron que el EDI-2 presentó adecuada consistencia interna y utilidad diagnóstica, por tanto puede utilizarse como un instrumento de tamizaje de los TCA en población mexicana y, dependiendo del objetivo del tamizaje, se puede seleccionar entre el punto de corte que presenta mayor sensibilidad o el que presenta mayor especificidad.

Cuestionario de Influencias sobre el Modelo Estético Corporal (CIMEC)

Objetivo: Evaluar las influencias culturales relevantes que contribuyen a promover y justificar el adelgazamiento, especialmente por razones estéticas y sociales.

Descripción: Incluye 40 reactivos con tres opciones de respuesta en escala tipo Likert, que se califican de 0 a 2.

El CIMEC fue creado y validado por Toro, Salameiro y Martínez (1994) en una muestra de 59 mujeres españolas diagnosticadas con AN, de acuerdo a los criterios del DSM-III-R (APA, 1987), y una muestra control ($n = 59$). La puntuación total y la de 26 de los 40 reactivos diferenciaron entre las muestras en la dirección esperada; por tanto, los autores calcularon la consistencia interna para los 40 reactivos ($Alpha = .94$ en la muestra con AN y .92 en la muestra control) y para los 26 reactivos (.93 y .91, respectivamente). El AFE se realizó sólo con los 26 reactivos y se obtuvieron cinco factores que explicaron 52.7% de la varianza: Malestar debido a la Imagen Corporal, Influencia de la Publicidad, Influencia de los Mensajes Verbales, Influencia de los Modelos Sociales e Influencia de las Situaciones Sociales. Asimismo, se propuso un punto de corte tanto para el CIMEC-40 como para el CIMEC-26, encontrándose que el primero mostró adecuada sensibilidad (81.4%) con la puntuación 23-24, pero deficiente especificidad (55.9%) y, para el CIMEC-26, el mejor punto de corte fue 16-17, con el que mejoraron ambos indicadores (83.1 y 64.4%, respectivamente), aunque la especificidad aún resultó poco satisfactoria. Sin embargo, debe considerarse que el propósito

de este cuestionario no es realizar un tamizaje de los posibles casos con TCA. Los autores concluyeron que el CIMEC es adecuado para evaluar la influencia de factores sociales en relación con el ideal de delgadez, mismos que promueven conductas alimentarias de riesgo, en su mayoría a través de los medios de comunicación.

La adaptación y validación del CIMEC-40 en mujeres mexicanas fue realizada por Vázquez, Alvarez y Mancilla (2000), en una muestra de mujeres estudiantes ($n = 1001$). La consistencia interna del instrumento fue $Alpha .94$ y se comprobó su homogeneidad, debido a que todos los reactivos correlacionaron significativamente con la puntuación total. El AFE derivó cuatro factores que agruparon 35 reactivos y explicaron 45.3% de la varianza total: Influencia de la Publicidad ($Alpha = .92$), Malestar por la Imagen Corporal ($Alpha = .86$), Influencia de los Modelos Estéticos Corporales ($Alpha = .75$), e Influencia de las Relaciones Sociales ($Alpha = .63$). Se concluyó que el CIMEC es confiable y válido para evaluar la influencia de los modelos estéticos corporales que se transmiten a través de la publicidad y las situaciones sociales.

Cuestionario de Tres Factores de la Alimentación (TFEQ, por sus siglas en inglés)

Objetivo: Evaluar tres dimensiones de la conducta alimentaria: control consciente, desinhibición y susceptibilidad al hambre.

Descripción: Está conformado por 51 reactivos que se dividen en dos apartados, 36 con un formato de respuesta dicotómico y 16 con cuatro opciones de respuesta en escala tipo Likert.

El TFEQ (Stunkard & Messick, 1985) fue construido a partir de la Revised Restraint Scale (Herman & Polivy, 1980), del Latent Obesity Questionnaire (Pudel, Metzendorff & Oetting, 1975) y la experiencia clínica de los autores. El TFEQ fue aplicado en 53 personas que hacían dieta y 45 personas que no hacían dieta, ambas muestras acudían a un programa para reducción de peso. Considerando la correlación reactivo-factor se seleccionaron 58 ítems que fueron

ingresados al AFE, el cual derivó tres factores que agruparon 55 reactivos: Restricción ($Alpha = .93$), Desinhibición ($Alpha = .91$) y Hambre ($Alpha = .85$). Sin embargo, al replicar el análisis en las dos muestras, esta estructura sólo se confirmó en las personas que hacían dieta, mientras que en la muestra de personas que no hacían dieta sólo se identificaron los factores Restricción y Hambre. El análisis del contenido de los reactivos resultó en la eliminación de cuatro ítems, por tanto la versión final del instrumento incluyó 51 reactivos. Los autores concluyeron que el TFEQ contribuye al estudio de la conducta de alimentación, ya que la identificación de las tres dimensiones permite estudiar aspectos específicos de ésta y sugieren que es necesario analizar la estabilidad del factor Desinhibición.

La versión en español del TFEQ (Sánchez-Carracedo, Raich, Figueras, Torras & Mora, 1999) fue adaptada y validada por López-Aguilar et al. (2011) en una muestra de 604 mujeres mexicanas (375 asistían a instituciones educativas, 143 a centros de ejercicio aeróbico-reductivo, 39 a tratamiento para TCA y 47 a solicitar evaluación por un posible TCA). La consistencia interna del instrumento fue $Alpha = .89$ y todos los reactivos correlacionaron de manera significativa con la puntuación total. El AFE derivó dos factores que agruparon 42 reactivos y explicaron 30% de la varianza total: Desinhibición ($Alpha = .91$) y Restricción Alimentaria ($Alpha = .84$). Los dos factores del TFEQ diferenciaron en la dirección esperada entre las mujeres con TCA ($n = 20$ con BN-Purgativa y 25 con TPA) y sin TCA ($n = 55$). Como evidencia de validez concurrente, se encontró que el factor Desinhibición correlacionó con el factor Bulimia del EAT-40 y Atracón del BULIT; en tanto que Restricción Alimentaria correlacionó con el factor correspondiente del EAT-40. Los resultados de la validación en México sugieren que el TFEQ presenta propiedades psicométricas adecuadas. No obstante, la estructura factorial, igual que en el estudio original de validación, mostró inestabilidad.

Hasta aquí se han expuesto las propiedades psicométricas de 11 instrumentos que fueron construidos en otros países, principalmente Estados Unidos, y

el análisis de sus propiedades psicométricas en población mexicana mostró que, en general, son herramientas confiables y válidas. Además, cabe señalar que la mayoría de los instrumentos son reconocidos en el contexto internacional por su utilidad como instrumentos de evaluación (cf. Anderson, Lundgren, Shapiro & Paulosky, 2004; Túry, Gülec & Kohls, 2010).

Cuestionarios desarrollados en México

Instrumentos de tamizaje

Cuestionario Breve de Conductas Alimentarias de Riesgo

Objetivo: Identificar conductas alimentarias de riesgo.

Descripción: Contiene 10 reactivos en escala tipo Likert con un formato de cuatro opciones de respuesta.

Este cuestionario fue creado por Unikel-Santoncini, Bojorquez-Chapela y Carreño-García (2004) para evaluar la preocupación por engordar, la práctica de atracones, la sensación de falta de control al comer y las conductas alimentarias de tipo restrictivo y purgativo, de acuerdo con los criterios diagnósticos del DSM-IV (APA, 1994). Las autoras trabajaron con una muestra de 78 pacientes diagnosticadas con TCA (48.7% BN, 9% AN-Restrictiva, 5.1% AN-Purgativa y 37.3% TCANE) y una muestra control de 468 estudiantes. La consistencia interna del instrumento fue $Alpha = .83$ y los 10 reactivos correlacionaron significativamente con la puntuación total. El AFE derivó tres factores que explicaron 64.7% de la varianza total: Atracón-Purga ($Alpha = .74$), Medidas Compensatorias ($Alpha = .72$) y Restricción ($Alpha = .76$). Al analizar la capacidad predictiva del instrumento, se propuso el punto de corte de 10 (sensibilidad = 81% y especificidad = 78%), en tanto que el análisis discriminante mostró que 88% de los casos fueron clasificados correctamente. Respecto a la validez concurrente se encontró que la puntuación total del cuestionario correlacionó con las escalas de sintomatología depresiva e ideación suicida ($r = .54$ y

.50, respectivamente) del Center of Epidemiological Studies of Depression (CES-D; Radloff, 1977; Roberts, 1980). Las autoras concluyeron que el cuestionario permite la identificación temprana de casos con TCA; además, resulta recomendable por ser una alternativa breve, de fácil aplicación y calificación.

Cuestionario Multidimensional de Factores de Riesgo Asociados a Trastornos de la Conducta Alimentaria

Objetivo: Medir conductas alimentarias y aspectos socioculturales asociados a los TCA.

Descripción: Existen dos versiones, una para varones y otra para mujeres. Dado el propósito de la presente revisión, a continuación sólo se hará referencia a la versión femenina, la cual está integrada por 23 reactivos con cuatro opciones de respuesta en escala tipo Likert, desde nunca hasta siempre.

Saucedo-Molina y Unikel-Santoncini (2010) desarrollaron este instrumento con base a la modificación e integración de dos cuestionarios: la Escala de Factores de Riesgo Asociados a los Trastornos Alimentarios y el CIMEC. El cuestionario resultante fue aplicado a una muestra de 632 mujeres adolescentes, encontrando que la consistencia interna fue adecuada ($Alpha = .83$). El AFE derivó cinco factores que explicaron 57.6% de la varianza: Influencia de la Publicidad ($Alpha = .86$), Desarrollo Puberal ($Alpha = .79$), Tiempos de Comida ($Alpha = .79$), Malestar con la Imagen Corporal ($Alpha = .72$) y Dieta Restrictiva ($Alpha = .71$). Las autoras concluyeron que el instrumento resulta válido para medir factores de riesgo asociados a TCA, principalmente aspectos socioculturales, conductas alimentarias y desarrollo puberal. Sin embargo, estos resultados sólo fundamentan la consistencia interna y la estructura factorial del cuestionario.

Instrumentos que miden factores asociados a los TCA

Instrumento de Imagen Corporal

Objetivo: Medir factores de riesgo asociados con

imagen corporal.

Descripción: Existe una versión de papel-lápiz y una versión computarizada. La primera incluye 30 reactivos (seis evalúan imagen corporal global y 24 partes corporales específicas). Contiene dos escalas visuales con nueve siluetas que se califican desde 1 (la más delgada) hasta 9 (la más gruesa), y un reactivo escalar de autopercepción del peso corporal con cinco opciones de respuesta, desde muy delgada hasta muy gruesa. La medición de partes corporales contiene ocho conceptos, cada uno con tres escalas tipo Osgood y cinco opciones de respuesta. En todos los reactivos, puntuaciones altas indican mayor problemática. La versión computarizada está conformada por 46 reactivos (seis miden imagen corporal, 16 miden partes corporales de frente y 24 imágenes de perfil), e incluye las nueve siluetas de frente utilizadas en la versión de papel-lápiz. Para la medición de partes corporales se presenta una silueta de frente dividida en ocho partes, que puede aumentar o disminuir de tamaño para construir la figura actual y la ideal; asimismo, se presenta una silueta de perfil dividida en seis partes que también pueden modificarse.

Este instrumento fue construido por Gómez-Peresmitré, Granados, Jáuregui, Pineda y Tafoya (2000), quienes aplicaron ambas versiones a 185 estudiantes universitarias que consideraban tener problemas con su peso corporal o con su forma de comer (muestra subclínica) y 127 estudiantes universitarias sin problemas con su peso o alimentación (muestra control). En ambas muestras, los coeficientes de consistencia interna fueron más altos para la versión computarizada ($Alpha = .90$ y $.94$, respectivamente) que para la versión de papel-lápiz ($Alpha = .80$ y $.84$, respectivamente). Las puntuaciones de las dos versiones del instrumento diferenciaron entre las muestras en la dirección esperada. Asimismo, se encontró que la versión computarizada clasificó correctamente a 98.5% de las participantes, en tanto que la otra versión a 87.3%. Se concluyó que las propiedades psicométricas de ambas versiones son apropiadas, aunque la versión computarizada mostró mejores resultados; además de que, dada su misma naturaleza, favorece un análisis más exhaustivo de los aspectos

de la imagen corporal.

Cuestionario de Actitudes hacia la Figura Corporal

Objetivo: Evaluar la internalización del ideal de delgadez.

Descripción: Está conformado por 15 reactivos en escala tipo Likert con cuatro opciones de respuesta, desde nunca hasta siempre.

Este cuestionario fue diseñado por Unikel, Juárez y Gómez (2006). Para la construcción de los reactivos se consideró la experiencia de mujeres con TCA y se retomaron cuatro ítems de un instrumento utilizado en México, sin embargo, los autores no especifican el nombre de dicho instrumento. El cuestionario fue validado en una muestra de 408 mujeres estudiantes y 87 pacientes en tratamiento para TCA. La consistencia interna fue $Alpha = .93$ y todos los reactivos correlacionaron con la puntuación total. El AFE derivó dos factores que explicaron 60% de la varianza total: Internalización ($Alpha = .90$) y Creencias ($Alpha = .88$). Esta estructura fue comprobada por medio de análisis factorial confirmatorio; además, se encontró que 83.7% de los casos fueron clasificados correctamente. La validez convergente fue corroborada al encontrar que la puntuación del cuestionario y sus dos factores correlacionaron significativamente con el Cuestionario Breve de Conductas Alimentarias de Riesgo (Unikel-Santoncini et al., 2004), la Escala de Sintomatología Depresiva del CES-D (Radloff, 1977), la subescala Auto-percepción de la Self-esteem Scale (Pope, McHale & Craighead, 1988) y la subescala Insatisfacción Corporal del EDI. Al analizar la predicción diagnóstica de la puntuación total del cuestionario, se encontró que el mejor punto de corte fue 37, con adecuada sensibilidad y especificidad (80%, en ambos casos). Los autores concluyeron que las propiedades psicométricas del instrumento fueron adecuadas.

Escala de Imagen Corporal en Mujeres Jóvenes

Objetivo: Medir imagen corporal.

Descripción: Consta de 26 reactivos con cinco op-

ciones de respuesta en escala tipo Likert, desde nunca (1) hasta siempre (5).

Esta escala fue creada por Rodríguez, Oudhof, González-Arratia y Unikel-Santoncini (2010), considerando tres indicadores de la imagen corporal: insatisfacción corporal, influencia social e interiorización del ideal estético de delgadez. La primera versión fue piloteada en 100 mujeres universitarias y se hicieron algunos ajustes en las instrucciones y en el orden de los reactivos. Posteriormente, para analizar la validez de contenido, se solicitó a cuatro jueces expertos que determinaran la pertinencia de los reactivos de acuerdo a los indicadores propuestos. La versión final de la escala quedó conformada por 33 reactivos, 18 fueron elaborados por los investigadores y 15 retomados del Cuestionario de Actitudes hacia la Figura Corporal (Unikel-Santoncini et al., 2006). La escala se aplicó en una muestra probabilística estratificada que incluyó 508 mujeres universitarias, muestra que resultó adecuada para el AFE ($KMO = .94, p < .001$), el cual confirmó la estructura factorial prevista por los autores, con base a 26 reactivos que explicaron 43.2% de la varianza total: Insatisfacción Corporal ($Alpha = .84$), Interiorización del Ideal de Delgadez ($Alpha = .89$) e Influencia Social ($Alpha = .82$). La consistencia interna para los 26 ítems agrupados en los factores fue $Alpha = .94$. Para examinar la validez predictiva de la escala se realizó un análisis discriminante, encontrándose que 13 reactivos clasificaron correctamente la imagen corporal en 98.2% de los casos con conductas alimentarias de riesgo y 97.4% de los casos agrupados mediante validación cruzada. También se encontró que los tres factores diferenciaron en la dirección esperada entre mujeres con conductas alimentarias de riesgo y mujeres sin conductas alimentarias de riesgo. La evidencia empírica obtenida en esta investigación indica que la escala presenta adecuada consistencia interna y resultados favorables respecto a su validez.

Discusión

En el año 2000 se publicaron los tres primeros trabajos en los que se validaron instrumentos dirigidos a

evaluar aspectos de los TCA en mujeres mexicanas: el BULIT (Alvarez et al. 2000), el Instrumento de Imagen Corporal (Gómez-Peresmitré et al., 2000) y el CIMEC (Vázquez et al., 2000). Éste fue el inicio de una fructífera producción científica en el campo de evaluación de los TCA. Actualmente se cuenta con 16 instrumentos validados en México, de los cuales cinco fueron creados por investigadores nacionales. Por tanto, cuando un profesional de la psicología o un investigador se disponga a realizar una evaluación sobre TCA, podrá elegir entre varias alternativas, dependiendo de su objetivo y de las garantías que el instrumento le ofrezca en términos de sus propiedades psicométricas.

La confiabilidad es uno de los aspectos que puede guiar la elección de un instrumento. Existen tres métodos para obtener estimaciones sobre la confiabilidad de una medida: el test-retest, el de formas paralelas y el método centrado en una sola aplicación (Nunally & Bernstein, 1995). Desde el punto de vista teórico, y si el instrumento así lo permite, la aplicación repetida del test en dos momentos temporales distintos es el método idóneo para evaluar la confiabilidad, particularmente en cuanto a la estabilidad de la medida sobre el tiempo (Muñiz, 1998). En las investigaciones revisadas, se observó que la tendencia general fue estimar la confiabilidad a partir del método de una sola aplicación, específicamente, calculando el coeficiente $Alpha$ de Cronbach, método basado en el análisis de la consistencia interna. Nunally y Bernstein (1995) sugieren que para decidir si la confiabilidad de un instrumento es satisfactoria, debe considerarse para qué se empleará. Por ejemplo, si el instrumento será utilizado con fines de investigación, un coeficiente de consistencia interna de .70 puede ser adecuado, pero si el instrumento se utilizará como auxiliar en la formulación de un diagnóstico o para tomar decisiones respecto a una persona, el valor mínimo debe situarse alrededor de .80. Tomando en cuenta esta guía, todos los instrumentos en los que se evaluó la consistencia interna resultan apropiados para los dos fines. Sin embargo, es importante mencionar que sólo en el caso del ChEAT (Escoto & Camacho, 2008) se evaluaron las dos propiedades de

la confiabilidad, tanto la consistencia interna como la estabilidad de la medida en el tiempo. Por tanto, es conveniente que investigaciones futuras examinen ambas propiedades, aportando mayor solidez empírica a la confiabilidad de las medidas.

El que un instrumento cuente con una confiabilidad alta no garantiza su validez, para ello deben considerarse las evidencias que específicamente corresponden a esta propiedad psicométrica. Para analizar los resultados sobre la validez de un instrumento es necesario tomar en cuenta que ésta comprende un proceso que comienza con la construcción del mismo y continúa, de manera indefinida, en la búsqueda de evidencias que enriquezcan el significado de las puntuaciones (Anastasi, 1986). La validez de un instrumento no puede resumirse en términos de un estadístico único y preciso. La determinación de la idoneidad, significado y utilidad del instrumento requiere la acumulación de evidencias sólidas que demuestren que cumple con el propósito para el que fue creado (Gregory, 2001). Los diferentes procedimientos para obtener evidencia sobre la validez se clasifican en tres categorías: validez de contenido, validez de constructo y validez de criterio.

Respecto a las investigaciones revisadas, se observa que los autores analizaron al menos un tipo de validez del instrumento en cuestión. Cuando se construye un instrumento, debe garantizarse la validez de contenido proporcionando evidencia de que los reactivos elaborados son relevantes y representan adecuadamente la definición semántica del constructo (Carretero-Dios & Pérez, 2005). Sin embargo, de los cinco instrumentos creados en México, sólo para la Escala de Imagen Corporal (Rodríguez et al., 2010) se informó explícitamente cómo se garantizó su validez de contenido; mientras que en tres, el Cuestionario Multidimensional de Factores de Riesgo asociados a Trastornos de la Conducta Alimentaria (Saucedo-Molina & Unikel-Santoncini, 2010), el Cuestionario Breve de Conductas Alimentarias de Riesgo (Unikel-Santoncini et al., 2004) y el Cuestionario de Actitudes hacia la Figura Corporal (Unikel-Santoncini et al., 2006), si bien los autores no hicieron explícita la información respecto a este tipo de validez, ésta

puede deducirse a partir de la lectura del apartado en el que describen cómo fue construido el instrumento. En tanto que en el caso del Instrumento de Imagen Corporal (Gómez-Peresmitré et al., 2000), no se informó cómo fue evaluada la validez de contenido.

La validez de constructo es quizás la más analizada y, generalmente, con base a la aplicación del AFE. Precisamente, este análisis fue utilizado en todos los estudios en donde se evaluó la estructura factorial de los instrumentos. Un requisito indispensable para llevar a cabo este análisis es que los reactivos se encuentren relacionados entre sí; es decir, se requiere que la matriz de correlaciones sea adecuada para poder identificar agrupamientos de reactivos. Para ello es necesario calcular ciertos estimadores, siendo las pruebas de elección la de esfericidad de *Bartlett* y el índice de *Kaiser-Meyer-Olkin*. En la presente revisión se encontró que sólo en la Escala de Imagen Corporal (Rodríguez et al., 2010), se informaron dichos estimadores. Asimismo, se ha señalado la conveniencia de que el AFE se replique en una segunda muestra (Carretero-Dios & Pérez, 2005), pero en ninguno de los estudios descritos se realizó, por lo que queda abierta esta línea de investigación. Esto es de particular relevancia si se considera que, por ejemplo, en el caso del EDI, la estructura factorial se analizó en dos estudios y mostró inestabilidad (Mancilla et al., 2003; Unikel, Bojorquez et al., 2006), sin embargo, los resultados no son directamente comparables; primero, porque se emplearon criterios diferentes para la retención de los reactivos (e.g. correlación reactivo-total, cargas factoriales) y, segundo, por el diferente tipo de muestra en que se analizó la estructura factorial, ya que Mancilla et al. lo hicieron en muestra comunitaria, mientras que Unikel et al. en mujeres con TCA. Por tanto, es necesario que en las investigaciones se procure emplear criterios homogéneos.

También es importante mencionar que en los instrumentos traducidos y/o adaptados, el AFE agrupó un menor número de reactivos de los que originalmente contiene el cuestionario, resultado que no se analizó en cuanto a sus posibles implicaciones. Este hecho sugiere que investigaciones posteriores debe-

rán contribuir a tomar la decisión de eliminar o conservar los reactivos de los instrumentos. Decisión que debe estar basada tanto en la valoración conjunta de todos los índices estadísticos posibles, como de los aspectos teórico-conceptuales que motivaron la creación del instrumento.

En algunas de las investigaciones revisadas también se presentó evidencia sobre la validez de criterio, en la que resulta de particular relevancia el análisis de la capacidad predictiva de los instrumentos, es decir, en qué medida la puntuación obtenida permite clasificar correctamente entre quienes tienen un TCA y quienes no lo tienen. En este sentido, la derivación del punto de corte de un instrumento es una aportación importante debido a que se generan herramientas de tamizaje o criba que permiten identificar a aquellas personas que están en una condición de mayor riesgo para el desarrollo de algún TCA. Específicamente, los instrumentos que cuentan con al menos un punto de corte para su empleo en mujeres mexicanas son: el EAT-40, el BULIT, el BITE, el BSQ, el EDI-2, el Cuestionario Breve de Conductas Alimentarias de Riesgo y el Cuestionario de Actitudes hacia la Figura Corporal. En general, el punto de corte propuesto para estos instrumentos mostró adecuada sensibilidad y especificidad (con porcentajes de predicción correcta ≥ 80), y sólo en el caso de dos los valores obtenidos fueron ligeramente menores; así, la sensibilidad del BULIT fue del 79% (Alvarez, 2000), mientras que la especificidad del Cuestionario Breve de Conductas Alimentarias de Riesgo fue del 78% (Unikel-Santoncini et al., 2004). Cabe señalar que en el caso de los estudios de validación del BULIT y del EDI-2 (García-García et al., 2006), el tamaño de la muestra clínica con que se estimó y examinó el punto de corte fue reducido (menos de 50 pacientes), por lo se requiere confirmar su pertinencia en una muestra mayor.

Se ha señalado que es deseable contar con las mayores evidencias posibles sobre la validez de los instrumentos, lo que se traduce en evaluar los tres tipos de validez. Sin embargo, para dos de los instrumentos analizados sólo se presentan resultados sobre la validez de constructo, tal es el caso del CIMEC

(Vázquez et al., 2000) y del BES (Zúñiga & Robles, 2006), por tanto, resulta necesario extender la investigación sobre la validez de estas medidas.

Por otro lado, es de notar que la mayoría de las investigaciones incluyeron muestras de la Ciudad de México y su zona metropolitana. Pero, considerando que en otros estados también se han identificado conductas alimentarias de riesgo, es necesario que futuras investigaciones incluyan muestras provenientes de diferentes regiones del país, para generar evidencias más sólidas respecto a las propiedades psicométricas de los instrumentos más allá de la capital del país.

Es necesario puntualizar que la presente revisión se limitó al análisis de los instrumentos validados en mujeres, principalmente como resultado de que la investigación nacional, seguramente como reflejo de la ruta de desarrollo de la realizada en el mundo, empezó enfocándose en muestras femeninas. No obstante, en las últimas décadas se ha reconocido que los TCA no son exclusivos del género femenino, por lo que será necesario analizar si los instrumentos creados considerando principalmente a este género, también son útiles para la evaluación de los varones. Y si bien, se tiene conocimiento de que los investigadores nacionales han empezado a trabajar en la creación, adaptación y evaluación de las propiedades psicométricas de herramientas que permitan examinar la imagen corporal y/o la conducta alimentaria en varones, aún son escasas las evidencias que se han publicado. Asimismo, se identificó la necesidad de trabajar sobre las formas de evaluación en niños, ya que hasta el momento sólo se ha validado el ChEAT (Escoto & Camacho, 2008).

La presente revisión evidencia que en el ámbito nacional existen avances importantes en el área de la evaluación de los TCA, independientemente del camino que se haya seguido, adaptar o crear el instrumento. No obstante, como resultado del balance realizado, también emergen vacíos y nuevos retos, delineando hacia dónde podrían dirigirse los actuales y futuros esfuerzos. Debe recordarse que la validación de una prueba es un proceso continuo que exige evidencias provenientes de diferentes fuentes

que puedan corroborar, complementar o extender los hallazgos existentes, buscando además fundamentar en qué medida un instrumento resulta o no vigente, en razón de que muestra ser útil y responder adecuadamente ante los cambios que se generan en la comprensión de la variable y/o constructo que miden. Por último, otro aspecto a considerar es que parece haber incrementado el conocimiento y el perfeccionamiento, por parte de los investigadores, en cuanto a la aplicación de técnicas estadísticas que permiten análisis más finos y precisos sobre las propiedades psicométricas de los instrumentos de medición, situación que sin duda redundará en una mayor proyección y alcance de la investigación generada en el ámbito nacional.

Agradecimientos: Los autores agradecen el financiamiento PROMEP/103.5/11/714 otorgado a la Dra. Karina Franco Paredes, así como el otorgado a la Dra. Georgina L. Alvarez Rayón como colaboradora del proyecto financiado por CONACyT (No. de registro: 131865).

Referencias

- Alvarez, G. (2000). *Validación en México de dos instrumentos para detectar trastornos alimentarios: EAT y BULIT*. Tesis de Maestría en Psicología Clínica, Facultad de Psicología. Universidad Nacional Autónoma de México.
- Alvarez, G., Mancilla, J. M., & Vázquez, R. (2000). Propiedades psicométricas del Test de Bulimia (BULIT). *Revista Psicología Contemporánea*, 7, 74-85.
- Alvarez-Rayón, G., Mancilla-Díaz, J. M., Vázquez-Arévalo, R., Unikel-Santoncini, C., Caballero-Romo, A. & Mercado-Corona, D. (2004). Validity of the Eating Attitudes Test: A study of Mexican patients with eating disorders. *Eating and Weight Disorders*, 9, 243-248.
- American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders (3rd ed., Text Revised)*. Washington, DC: Author.
- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders (4th ed.)*. Washington, DC: Author.
- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders (4th ed., Text Revised)*. Washington, DC: Author.
- Anastasi, A. (1986). Emerging concepts of test validation. *Annual Review of Psychology*, 37, 1-15.
- Anderson, D. A., Lundgren, J. D., Shapiro, J. R. & Paulosky, C. A. (2004). Assessment of eating disorders: Review and recommendations for clinical use. *Behavior Modification*, 28, 763-782.
- Carretero-Dios, H. & Pérez, C. (2005). Normas para el desarrollo y revisión de estudios instrumentales. *International Journal of Clinical and Health Psychology*, 5, 521-551.
- Castro, J., Toro, J., Salamero, M. & Guimerá, E. (1991). The Eating Attitudes Test: Validation of the Spanish version. *Psychological Assessment*, 2, 175-190.
- Cooper, P., Taylor, M., Cooper, Z. & Fairburn, C. (1987). The development and validation of the Body Shape Questionnaire. *International Journal of Eating Disorders*, 6, 485-495.
- Coopersmith, S. (1967). *The antecedents of self-esteem: A psychiatric view*. San Francisco, CA: Consulting Psychologist Press.
- Derogatis, L. R., Lipman, R. S. & Covi, L. (1973). SCL-90: An outpatient psychiatric rating scale-preliminary report. *Psychopharmacological Report*, 9, 13-27.
- Escoto, M. C. & Camacho, E. J. (2008). Propiedades psicométricas del Test Infantil de Actitudes Alimentarias en una muestra mexicana. *Revista Mexicana de Psicología*, 25, 99-106.
- Feighner, J. P., Robins, E., Guze, S., Woodruff, R., Winokur, G. & Muñoz, R. (1972). Diagnostic criteria for use in psychiatric research. *Archives of General Psychiatry*, 26, 57-63.
- Franco, K., González, O. L., Díaz, F. J., López-Espinoza, A., Martínez, A. G. & Aguilera, V. (2010). Reliability and validity of Bulimic Investigatory Test Edinburgh on Mexican women. *Journal of Behavior, Health & Social Issues*, 2, 17-24.
- García-García, E., Vázquez-Velázquez, V., López-Alvarenga, J. & Arcila-Martínez, D. (2003). Validez interna y utilidad diagnóstica del Eating Disorders Inventory en mujeres mexicanas. *Salud Pública de México*, 45, 206-210.
- Garner, D. M. (1991). *Eating Disorders Inventory-2: Professional manual*. Odessa, FL: Psychological Assessment Resources.
- 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., Bohr, Y. & Garfinkel, P. E. (1982). The Eating Attitudes Test: Psychometric features and clinical correlates. *Psychological Medicine*, 12, 871-878.
- Garner, D. M., Olmsted, M. P. & Polivy, J. (1983). Development and validation of a multidimensional Eating Disorder Inventory for anorexia nervosa and bulimia. *International Journal of Eating Disorders*, 2, 15-34.
- Gómez-Peresmitré, G., Granados, A., Jáuregui, J., Pineda, G. & Tafoya, A. (2000). Un instrumento para medir imagen corporal: Versión computarizada y de papel y lápiz. *Revista Mexicana de Psicología*, 17, 89-99.
- Gormally, J., Black, S., Daston, S. & Rardin, D. (1982). The assessment of binge eating severity among obese persons. *Addictive Behaviors*, 7, 47-55.
- Gregory, R. J. (2001). Evaluación psicológica: Historia, principios y aplicaciones. México: Manual Moderno.
- Guimerá, E. & Torrubia, R. (1987). Adaptación española del Eating Disorder Inventory (EDI): En una muestra de pacientes anoréxicas. *Anales de Psiquiatría*, 3, 185-190.
- Henderson, M. & Freeman, P. L. (1987). A self-rating scale for bulimia: The BITE. *British Journal of Psychiatry*, 150, 18-24.
- Herman, C. P. & Polivy, J. (1980). Restrained eating. En A.J. Stunkard (Ed.). *Obesity*. Philadelphia: Saunders.
- Kutlesic, V., Williamson, D., Gleaves, D., Barbin, J. & Murphy-Eberenz, K. (1998). The interview for diagnosis of eating disorders-IV: Application to DSM-IV diagnostic criteria. *Psychological Assessment*, 10, 41-48.
- López, X., Mancilla, J. M., Vázquez, R., Alvarez, G. & Franco, K. (2011). Un estudio exploratorio sobre propiedades psicométricas del Cuestionario de Patrones de Alimentación y Peso-Revisado (QEWPR). *Revista Psicología.com*, 1-11.
- López-Aguilar, X., Mancilla-Díaz, J. M., Vázquez-Arévalo, R., Franco-Paredes, K., Alvarez-Rayón, G. & Ocampo, M. T. (2011). Propiedades psicométricas del Cuestionario de Tres Factores de la Alimentación (TFEQ). *Revista Mexicana de Trastornos Alimentarios*, 2, 24-32.
- Maloney, M., Bell, J. & Daniels, S. (1988). Reliability testing of a children's version of the Eating Attitudes Test. *Journal of the American Academy of Child and Adolescents Psychiatry*, 5, 541-543.
- Mancilla, J. M., Franco, K., Alvarez, G. & Vázquez, R. (2003). Evaluation of the psychometric properties of the Mexican version of the Eating Disorders Inventory. *Revista Thomson Psicología*, 1, 167-176.
- Mancilla, J. M., Gómez-Peresmitré, G., Alvarez, G., Franco, K., Vázquez, R., López, X. et al. (2006). Trastornos del comportamiento alimentario en México. En J. M. Mancilla & G. Gómez-Peresmitre (Eds.). *Trastornos del comportamiento alimentario en Hispanoamérica* (pp. 123-171). México: Manual Moderno.
- Mora, M. (1992). *Bulimia y variables de riesgo: Un análisis causal*. España: Universidad Autónoma de Barcelona.
- Muñiz, J. (1998). *Teoría clásica de los tests*. Madrid: Pirámide.
- Nunnally, J. C. & Bernstein, I. J. (1995). *Teoría psicométrica*. Madrid: McGraw-Hill.
- Pope, A. W., McHale, S. M. & Craighead, W. E. (1988). *Self-esteem enhancement with children and adolescents*. USA: Allyn & Bacon.
- Pudel, V. E., Metzdorff, M. & Oetting, M. X. (1975). Zur persöhnlichkeit adipoeser in psychologischen tests unter beruecksichtigung latent fettuechtiger. *Zeitschrift für Psychosomatische Medizin und Psychoanalyse*, 21, 345-350.
- Radloff, L. (1977). The CES-D Scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*, 1, 385-401.
- Raich, R. M., Mora, M., Soler, A., Ávila, C., Clos, I. & Zapater, L. (1996). Adaptación de un instrumento de evaluación de la insatisfacción corporal. *Clínica y Salud*, 7, 51-66.
- Rivas, T., Bersabé, R. & Jiménez, M. (2004). Fiabilidad y validez del Test de Investigación Bulímica de Edimburgo (BITE) en una muestra de adolescentes españoles. *Psicología Conductual*, 12, 447-461.
- Roberts, E. (1980). The reliability of the CES-D Scale in different ethnic contexts. *Psychiatry Research*, 2, 125-134.
- Rodríguez, B., Oudhof, H., Gonzalez-Arratia, N. I. & Unikel-Santoncini, C. (2010). Desarrollo y validación de una escala para medir imagen corporal en mujeres jóvenes. *Salud Mental*, 33, 325-332.
- Sánchez-Carracedo, D., Raich, R. M., Figueras, M., Torras, J. & Mora, M. (1999). Adaptación preliminar del cuestionario de alimentación de Stunkard y Messick (Three Factor Eating Questionnaire, TFEQ) con una muestra española universitaria. *Psicología Conductual*, 7, 393-413.

- Saucedo-Molina, T. J. & Unikel-Santoncini, C. (2010). Validez de un instrumento multidimensional para medir factores de riesgo asociados a trastornos de la conducta alimentaria en púberes mexicanos. *Revista Chilena de Nutrición*, 37, 60-69.
- Smith, M. C. & Thelen, M. H. (1984). Development and validation of a test for bulimia. *Journal of Consulting and Clinical Psychology*, 52, 863-872.
- Spitzer, R. L., Devlin, M., Walsh, B., Hasin, D., Wing, R., Marcus, M. D. et al. (1992). Binge eating disorder: A multisite field trial of the diagnostic criteria. *International Journal of Eating Disorders*, 11, 191-203.
- Spitzer, R. L., Yanovski, S., Wadden, T., Wing, R., Marcus, M. D., Stunkard, A. J. et al. (1993). Binge eating disorder: Its further validation in a multisite study. *International Journal of Eating Disorders*, 13, 137-153.
- Stunkard, A. J. & Messick, S. (1985). The Three-Factor Eating Questionnaire to measure dietary restraint, disinhibition and hunger. *Journal of Psychosomatic Research*, 29, 71-83.
- Toro, J., Salamero, M. & Martínez, E. (1994). Assessment of sociocultural influences on the aesthetic body shape model in anorexia nervosa. *Acta Psychiatrica Scandinavica*, 89, 147-151.
- Túry, F., Gülec, H. & Kohls, E. (2010). Assessment methods for eating disorders and body image disorders. *Journal of Psychosomatic Research*, 69, 601-611.
- Unikel, C. & Bojorquez, I. (2007). A review of eating disorders research in Mexico. *International Journal of Psychology*, 42, 59-68
- Unikel, C., Bojorquez, I., Carreño, S. & Caballero, A. (2006). Validación del Eating Disorder Inventory en una muestra de mujeres mexicanas con trastorno de la conducta alimentaria. *Salud Mental*, 29, 44-51.
- Unikel, C., Juárez, F. & Gómez, G. (2006). Psychometric properties of the Attitudes Towards Body Figure Questionnaire in Mexican female students and patients with eating disorders. *European Eating Disorders Review*, 14, 430-435
- Unikel-Santoncini, C., Bojorquez-Chapela, I. & Carreño-García, S. (2004). Validación de un cuestionario breve para medir conductas alimentarias de riesgo. *Salud Pública de México*, 46, 509-515.
- Vázquez, R., Alvarez, G. & Mancilla, J. M. (2000). Consistencia interna y estructura factorial del Cuestionario de Influencias sobre los Modelos Estéticos Corporales (CIMEC) en población mexicana. *Salud Mental*, 23, 18-24.
- Vázquez, R., Galán, J., Aguilar, X., Alvarez, G., Mancilla, J. M., Caballero, A. et al. (2011). Validez del Body Shape Questionnaire (BSQ) en mujeres mexicanas. *Revista Mexicana de Trastornos Alimentarios*, 2, 42-52.
- Yela, M. (1997). *La técnica del análisis factorial*. Madrid: Biblioteca Nueva.
- Zúñiga, O. & Robles, R. (2006). Validez de constructo y consistencia interna del Cuestionario de Trastorno por Atracón en población mexicana con obesidad. *Psiquis*, 15, 126-134.