



Facultad de Estudios Superiores
IZTACALA

Psychometric properties of self-reports of emotional eating: A systematic review

Propiedades psicométricas de los autoreportes de ingesta emocional: Una revisión sistemática

Vanessa Monserrat Vázquez Vázquez¹, Karina Franco-Paredes¹, Felipe de Jesús Díaz-Reséndiz¹, Lilián Elizabeth Bosques-Brugada², Erin Yaqueline Tovar-Garza¹

¹ Universidad de Guadalajara, Centro Universitario del Sur

² Universidad Autónoma del Estado de Hidalgo, Instituto de Ciencias de la Salud

Recibido: 2023-05-18

Revisado: 2023-06-17

Aceptado: 2023-08-22

Autores de correspondencia: vanessa.vazquez4619@alumnos.udg.mx (V. M. Vázquez-Vázquez)

Financiamiento y Agradecimientos: Al Programa de Becas de Posgrado del CONAHCYT por el apoyo otorgado a la primera autora.

Conflicto de intereses: Los autores declaran no tener conflicto de intereses.

Resumen. El interés por la ingesta emocional (IE) ha ido en aumento, así como el número de autoreportes que permiten evaluarla. El objetivo de la presente investigación fue realizar una revisión sistemática de la calidad psicométrica de los autoreportes de IE en población adulta (PROSPERO ID: CRD42022374665). La búsqueda se realizó en *PubMed*, *Web of Science*, *Scopus*, *Springer link*, *Taylor & Francis*, *Science Direct* e identificación manual considerando los lineamientos PRISMA. La calidad metodológica y de las propiedades de medida fueron evaluadas a partir de los estándares y criterios COSMIN. Fueron analizados nueve autoreportes que denotaron discrepancias conceptuales y que en su mayoría se trataron de escalas multidimensionales. Las principales evidencias de validez reportadas fueron la estructural (a través de análisis factorial exploratorio o análisis de componentes principales) y de constructo, en menor medida, algunos artículos realizaron procedimientos cualitativos como parte de la validez de contenido. En tanto a la confiabilidad, el procedimiento más empleado fue el análisis de consistencia interna ($\alpha = .77 - .94$) y en menor medida la estabilidad temporal. La *Positive-Negative Emotional Eating Scale* y la *Salzburg Emotional Eating Scale* obtuvieron mejores calificaciones respecto a

su calidad metodológica y métrica. Las mayores deficiencias identificadas en los autoreportes fueron respecto a la validez de contenido, lo que pone en evidencia la necesidad de fortalecer esta propiedad métrica, así como la validez estructural y la consideración de otras poco o nulamente reportadas.

Palabras claves: revisión sistemática; autoreportes; ingesta emocional; psicometría; adultos.

Abstract. Interest in emotional eating (EE) has been increasing, as well as the number of self-report instruments that allow it to be evaluated. This study aimed to carry out a systematic review of the psychometric quality of EE self-reports in adult population (PROSPERO ID: CRD42022374665). The search was carried out in PubMed, Web of Science, Scopus, Springer link, Taylor & Francis, and Science Direct databases, and by manual identification considering PRISMA guidelines. Methodological quality and measurement properties were evaluated based on the COSMIN standards and criteria. Nine self-reports were analyzed, which denoted conceptual discrepancies and most of them were multidimensional scales. The main validity evidences reported were structural (through exploratory factor analysis or principal component analysis) and construct validity; to a lesser extent, some articles performed qualitative procedures as part of content validity. As for reliability, the most used procedure was the analysis of internal consistency ($\alpha = .77 - .94$) and to a lesser extent temporal stability. The Positive-Negative Emotional Eating Scale and Salzburg Emotional Eating Scale obtained better marks regarding their methodological and metric quality. The greatest deficiencies identified in the self-reports were regarding content validity, which highlights the need to strengthen this metric property, as well as structural validity and consideration of others that have been little or not reported at all.

Keywords: systematic review; self-reports; emotional eating; psychometrics; adults.

Introducción

La ingesta emocional (IE) entendida inicialmente como una estrategia de afrontamiento ante experiencias emocionales desagradables (Macht & Simons, 2011), empezó a estudiarse en personas con obesidad para analizar la relación de la sobreingesta alimentaria y los estados emocionales displacenteros. Estos mismos autores mencionan que la IE puede clasificarse en tres tipos: ocasional, habitual y atracón/compulsiva. En el primer tipo se consumen de manera ocasional pequeñas cantidades de alimentos como *snacks* para mejorar el estado de ánimo, mismo en el que interviene un mecanismo hedónico. En el segundo tipo, las personas habitualmente ingieren una comida completa para regular su estado de ánimo, están involucrados mecanismos fisiológicos y hedónicos. Mientras en el último tipo se presentan episodios de atracón, caracterizados por la ingesta compulsiva de

grandes cantidades de alimentos en los que se pueden tener efectos neuroquímicos o neuroendócrinos.

El principal problema de salud abordado con respecto a la IE es el exceso de peso y las dificultades en la pérdida del mismo en las intervenciones médicas. Frayn y Knäuper (2017) realizaron una revisión de la literatura para analizar la relación entre IE y exceso de peso. Encontraron que cinco estudios longitudinales reportaron consistentemente una asociación positiva entre la IE y el aumento de peso. Mientras la relación de la IE con la pérdida de peso en intervenciones médicas fue analizada en tres estudios, encontrándose que mayores puntajes de IE predecían una menor pérdida de peso, cabe aclarar que en dos estudios este hallazgo fue reportado únicamente en mediciones iniciales y no a largo plazo. Sin embargo, las autoras mencionan que los estudios analizados muestran heterogeneidad entre los tamaños del efecto y en los coeficientes de regresión subrayando la necesidad de estudios que

permitan lidiar con estas limitaciones y mejoren la calidad de la metodología. De manera similar, Smith et al. (2023) realizaron una revisión sistemática con meta-análisis de 34 artículos para analizar la efectividad de las intervenciones de IE en la pérdida de peso de adultos con sobrepeso u obesidad. Este estudio halló un efecto pequeño pero significativo de dichas intervenciones en el porcentaje del cambio de peso (-1.08% [IC 95 % = -1.66 a -0.49]) con una heterogeneidad considerable en los estudios ($I^2 = 64.65$ %). Lo anterior, pone en evidencia la relevancia de intervenir en la IE de pacientes con exceso de peso.

Además, en la población adulta la IE ante emociones desagradables y ambiguas (depresión, ansiedad, enojo y aburrimiento) está relacionada con un pobre bienestar psicológico, así como mayores síntomas de trastornos alimentarios y dificultades en la regulación emocional (Braden et al., 2018). En jóvenes de pregrado (mujeres y varones) la IE se ha relacionado con una mayor sintomatología depresiva, menores puntajes de salud autopercebida y de satisfacción con la vida, así como un mayor índice de masa corporal (Sze et al., 2021).

El *Ducht Eating Behavior Questionnaire* (DEBQ, por sus siglas en inglés; van Strien et al., 1986) fue el primer autoreporte que incluyó la medición de la IE como uno patrón alimentario asociado a la obesidad, este instrumento fue diseñado para población adulta. Sin embargo, la *Emotional Eating Scale* (EES-A, por sus siglas en inglés; Arnou et al., 1995) fue la primera escala en medir únicamente la IE, considerando tres dimensiones: Enojo/Frustración, Ansiedad y Depresión; también en adultos. En el DEBQ como en la EES-A, la IE es evaluada a partir de preguntar por modificaciones en el apetito a razón de experiencias emocionales displacenteras o ambiguas. Posterior a la construcción de la EES-A se ha evidenciado un creciente número de autoreportes que miden la IE. En esta variedad de instrumentos de IE se ha señalado la presencia de discrepancias entre las emociones consideradas y el comportamiento evaluado, mismas que podrían llevar a la obtención de resultados inconsistentes (Reichenberger et al., 2020).

A la fecha no se cuenta con algún estudio en el que se halla evaluado la calidad métrica de los instrumentos que miden la IE. Sin embargo, algunos trabajos han dado cuenta de la necesidad de verificar

las propiedades de medida de estos instrumentos. Bongers y Jansen (2016) realizaron una revisión de estudios tanto experimentales como ecológicos e identificaron deficiencias en la capacidad predictiva y discriminante de algunos de los autoreportes más usados para medir IE, entre ellos: el DEBQ (van Strien et al., 1986) desarrollado en una muestra de varones y mujeres de Países Bajos, la versión revisada de la EES-A (Koball et al., 2012) que tuvo como muestra a jóvenes de pregrado adscritos a una institución estadounidense y el Cuestionario de Three-Factor Eating Questionnaire-R (TFEQ-R, por sus siglas en inglés; Ganley, 1988) originalmente aplicado a mujeres adultas caucásicas. Asimismo, se ha subrayado la necesidad de elegir los instrumentos de IE bajo previo conocimiento de las definiciones conceptuales y operacionales de los mismos (Autor, 2019).

Si bien los autoreportes no son las únicas formas para evaluar la IE es sabido que son herramientas ampliamente utilizadas tanto para la investigación científica como en la práctica clínica (Autor, 2011). La construcción de un instrumento debe seguir un procedimiento riguroso y objetivo, considerando estándares para garantizar la calidad de las inferencias que se realizarán a partir de las puntuaciones obtenidas (Lane et al., 2016). Por ello, es importante que los autoreportes sean evaluados respecto a su calidad metodológica (riesgo de sesgo) y métrica. Estas evaluaciones permiten a los profesionales realizar elecciones adecuadas para los fines con los que pretenden utilizar los instrumentos.

Considerando la importancia que tienen los instrumentos psicológicos para el estudio de diferentes constructos, se han desarrollado lineamientos para evaluar su calidad métrica, entre estos los *Consensus-based Standards for the Selection of Health Measurement Instruments* —COSMIN— (Prinsen et al., 2018), que incluyen una serie de estándares y criterios para evaluar la calidad metodológica y la calidad de las propiedades métricas en revisiones sistemáticas.

Hasta la fecha se han construido diversos instrumentos para medir la IE, por lo que resulta importante facilitar la toma de decisiones instrumentales para investigadores y clínicos analizando con estándares internacionales las propiedades métricas de las escalas (Bongers & Jansen, 2016; Mookink et al.,

2016). Actualmente no se cuenta con un documento integrador del análisis de la calidad métrica de los autoreportes que exclusivamente evalúan la IE. El objetivo de la presente investigación fue realizar una revisión sistemática de la calidad psicométrica de los autoreportes de IE en población adulta. Para realizar el análisis se consideraron los estándares y criterios COSMIN (Prinsen et al., 2018).

Método

La presente revisión sistemática está registrada en PROSPERO (CRD42022374665) y se realizó de acuerdo con las directrices y lineamientos de la declaración *Preferred Reporting Items for Systematic Reviews and Meta-Analyses* —PRISMA— (Page et al., 2021).

Fuentes de información y estrategia de búsqueda

La temporalidad considerada en la búsqueda de instrumentos fue desde el año 1983 (primera publicación sobre IE) hasta el 22 de septiembre del 2022. La búsqueda se realizó en seis bases de datos: *PubMed/Medline*, *Web of Science*, *Scopus*, *Springer Link*, *Taylor y Francis* y *Science Direct*, además se realizó una búsqueda manual (esta última a partir de la identificación de referencias en los artículos incluidos), actividad desempeñada por la investigadora VMVV.

Para facilitar la identificación de artículos la formulación de la estrategia de búsqueda se consideró el acrónimo POT (*population, outcome and type of study*); P = adultos, O = IE y T = instrumental. El algoritmo de búsqueda fue el siguiente: ([Adults] AND ["Emotional eating" OR "Emotional eater" OR "Emotional overeating"]) AND [Instrumental OR Questionnaire OR Measure OR Self-report OR Scale OR "Psychometric properties" OR Validity OR Reliability OR Development OR Construction]). Únicamente en *Science Direct* se utilizó una fórmula reducida ([Adults] AND ["Emotional eating"] AND [Instrumental OR Questionnaire OR Measure OR Self-report OR Scale OR "Psychometric properties" OR Development]) debido a limitaciones en la cantidad de operadores booleanos permitidos en esta base de datos.

Criterios de selección de los estudios

La elegibilidad de los artículos se consideró en función del cumplimiento de los siguientes criterios de inclusión: 1) autoreportes originales que miden la IE; 3) idioma (inglés o español); 4) tipo de publicación (artículos científicos) y 5) población (adulta ≥ 18 años). Se excluyeron aquellos estudios que se trataron de: 1) documento no instrumental, 2) instrumentos que no evaluaran la IE o lo hicieran a través de una subescala que formara parte de un instrumento y 3) validaciones y/o revisiones de propiedades psicométricas de algún instrumento preexistente.

Procedimiento

Selección de los autoreportes

El proceso de selección de los estudios encontrados en las bases de datos se realizó en tres pasos: 1) exportación de las referencias y eliminación de documentos duplicados, paso realizado por VMVV; 2) revisión de título y resumen de los estudios (bajo la consideración de los criterios de selección) desde la plataforma de *Rayyan* (Ouzzani et al., 2016), actividad desempeñada por VMVV y FJDR —de forma independiente KFP resolvió las discrepancias surgidas—; 3) recuperación y análisis de los artículos a texto completo llevado a cabo por VMVV y KFP —EYTG y FJDR resolvieron las discrepancias surgidas—.

Análisis de la información

La evaluación de la calidad metodológica se realizó a través de los estándares COSMIN (Mokkink et al., 2018; Prinsen et al., 2018), utilizando un sistema de calificación que va de "muy bueno" (V) a "inadecuado" (I). Mientras que la calidad métrica de los artículos se evaluó con los criterios de las buenas propiedades métricas establecidas por COSMIN (Prinsen et al., 2018). Del grado de cumplimiento de cada criterio puede obtenerse una calificación positiva/suficiente (+), una calificación indeterminada (?) o una calificación insuficiente (-). Para ambas evaluaciones la puntuación

general se obtuvo por el método de la “peor puntuación obtenida”. La evaluación de la calidad metodológica y métrica fue realizada por VMVV y KFP. Se calculó el índice de acuerdo (coeficiente de concordancia W de Kendall). Las discrepancias fueron resueltas por mutuo acuerdo entre las autoras.

Resultados

Identificación y selección de instrumentos

El diagrama de flujo PRISMA (Figura 1) muestra el proceso de identificación y selección de instrumentos. Se observa que de 4,949 registros; 1,410 se trataron de artículos duplicados, por lo que fue revisado el título y resumen de 3,539 estudios, de los cuales 14 fueron leídos a texto completo y de ellos, siete cumplieron con los criterios de elegibilidad, asimismo se les sumaron dos artículos identificados por búsqueda manual en las referencias de los documentos previamente incluidos y que cumplieron con los criterios de selección, por lo que en total se analizaron nueve autoreportes.

Descripción general de los autoreportes

La descripción general de los autoreportes analizados se muestra en la Tabla 1. Se analizaron los siguientes instrumentos: EES-A (Arnold et al., 1995), el *Emotional Appetite Questionnaire* —EMAQ, por sus siglas en inglés— (Geliebter y Aversa, 2003), el *Emotional Overeating Questionnaire* —EOQ, por sus siglas en inglés— (Masheb & Grilo, 2006), la *Emotional Eating Scale* —EES-D, por sus siglas en inglés— (Doğan et al., 2011), el Cuestionario de Comedor Emocional —CCE, por sus siglas en español— (Garaulet et al., 2012), la Escala de Alimentación Emocional —EAE, por sus siglas en español— (Rojas & García-Méndez, 2017), la *Positive-Negative Emotional Eating Scale* —PNEES, por sus siglas en inglés— (Sultson et al., 2017), la *Salzburg Emotional Eating Scale* —SEES, por sus siglas en inglés— (Meule et al., 2018) y el *Florence Emotional Eating Drive* —FEED, por sus siglas en inglés— (Cassioli et al., 2021).

Tres de los instrumentos son de origen estadounidense y el resto pertenece a diferentes países. El número de reactivos que integran los autoreportes oscilan entre 6 y 36 reactivos ($M = 19.5$ y $Md = 20$) con una escala de respuesta tipo Likert, siendo de cinco opciones de respuesta la más empleada.

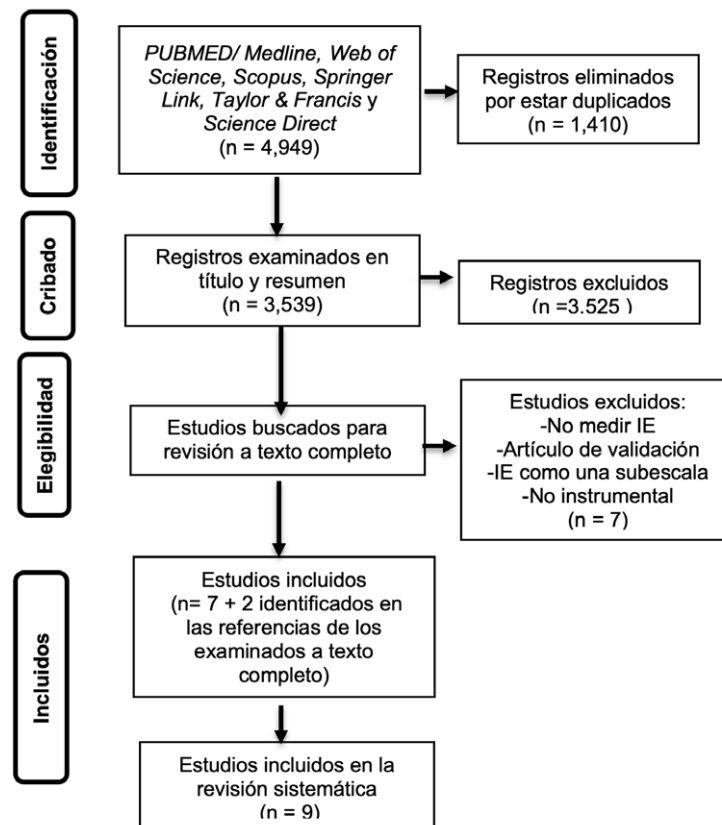
La mayoría de autoreportes conceptualizan la IE como el incremento o modificación de la conducta alimentaria debido a la experimentación de emociones displacenteras, no obstante, algunos de ellos han incluido también a las emociones positivas, señalando a la IE como una respuesta a los acontecimientos positivos y negativos para evitar, regular, enfrentar o mejorar la experiencia (Geliebter & Aversa, 2003; Meule et al., 2018; Rojas & García-Méndez, 2017; Sultson et al., 2017). Otra discrepancia en los autoreportes analizados está en el tipo de información indagada (modificación del apetito, frecuencia de la ingesta o la cantidad consumida).

Siete de los nueve autoreportes son multidimensionales: la EES-A (Arnold et al., 1995), el EMAQ (Geliebter & Aversa, 2003), el CCE (Garaulet et al., 2012), la EAE (Rojas & García-Méndez, 2017), la PNEES (Sultson et al., 2017), la SEES (Meule et al., 2018) y el FEED (Cassioli et al., 2021); de los cuales, cuatro de ellos tienen como dimensiones estados emocionales; la EES-A (Arnold et al., 1995) y el FEED (Cassioli et al., 2021): Enojo, Ansiedad y Depresión. La SEES (Meule et al., 2018): Tristeza, Enojo, Ansiedad y Felicidad. Mientras la PNEES (Sultson et al., 2017) se compone de la dimensión Ingesta Emocional Positiva e Ingesta Emocional Negativa.

El EMAQ (Geliebter & Aversa, 2003) incluye dos dimensiones de estados emocionales (una positiva y otra negativa) así como una de situaciones “positivas” y otra de situaciones “negativas”. En el caso del CCE (Garaulet et al., 2012) sus dimensiones están más enfocadas en la IE en episodios de atracón (Desinhibición, Tipo de Comida y Culpa). El EOQ (Masheb & Grilo, 2006) y la EES-D (Doğan et al., 2011) se tratan de autoreportes unidimensionales. En el caso del EOQ (Masheb & Grilo, 2006) está constituido por seis reactivos; cinco emociones negativas (Ansiedad, Tristeza, Soledad, Cansancio e Ira) y una de valencia positiva (Felicidad). Mientras que la EES-D (Doğan et al., 2011) incluye 14 emociones de valencia negativa (aunque estas no se especifican).

El tamaño de la muestra total reportado en los artículos osciló entre 98 – 1,698 ($M = 549.11$ y $Md = 420$)

Figura 1. Proceso de identificación y selección de los instrumentos de IE (diagrama de flujo PRISMA)



participantes. Siete estudios consideraron participantes de ambos sexos (varones y mujeres) con una proporción mayor de mujeres ($M = 67.51\%$, $Md = 68.81\%$); mientras que los dos artículos restantes sólo incluyeron mujeres. El rango de edad de los y las participantes fue de 15 a 72 años ($M = 33.18$, $Md = 33.5$), con una mayor participación de personas en etapa de adultez temprana.

En cuatro de los nueve artículos se especifica la categoría del peso a la que pertenecían los participantes, en tres de ellos eran personas con obesidad, mientras que en el estudio del EMAQ (Geliebter & Aversa, 2003) se especifica haber incluido a personas con un IMC perteneciente al bajo peso, normopeso y sobrepeso.

Evaluación de la calidad metodológica y métrica

En la Tabla 2 se observa la evaluación de la calidad metodológica y de la calidad métrica de acuerdo con los

estándares y criterios COSMIN (Mokkink et al., 2018; Prinsen et al., 2018). El coeficiente de concordancia W de Kendall fue de .88 ($p < .001$).

Validez de contenido

De los nueve artículos evaluados únicamente cuatro de ellos mencionaron haber realizado procedimientos cualitativos para revisar la validez de contenido, ya sea a través de entrevistas a personas de la población de interés o la consulta con expertos. Este es el caso de la EES-A (Arnold et al., 1995), la EES-D (Doğan et al., 2011), la EAE (Rojas & García-Méndez, 2017) y el FEED (Cassoli et al., 2021). De las cuales, la EES-D (Doğan et al., 2011) informó una mayor cantidad de procedimientos para analizar la validez de contenido, sin embargo, no se cumplieron los estándares de calidad metodológica en su totalidad, por lo que obtuvo una calificación dudosa y el resto de los instrumentos se

Tabla 1. Datos generales de los autoreportes de IE

Instrumento (autor, año)	País	Ítems	Escala Likert	Factores	Muestra		
					n(% de mujeres)	(M, DE, rango de años)	Característica
EES-A (Arnou et al., 1995)	EE.UU.	26	1 “no deseo comer” a 5 “una necesidad imperante de comer”	1) Enojo, 2) Ansiedad y 3) Depresión	E1: 47(100%) E2: 51(100%) nT-r = E1	E1: M = 44.9, DE = 10.4, 23-64 E2: M = 45.1, DE = 10.6, 21-65	Adultos con obesidad
EMAQ (Gelibter & Aversa, 2003)	EE.UU.	22	1 “mucho menos” a 9 “mucho más”, N/A y DK	1) EMAQ-PE 2) EMAQ-PS 3) EMAQ-NE 4) EMAQ-NS	90(50%) nT-r: 90(50%)	GB: M = 29.2, DE = 9.6, NM GN: M = 28.9, DE = 6.6, NM GS: M = 33.5, DE = 11.2, NM	Adultos con diferentes IMC
EOQ (Masheb & Grilo, 2006)	EE.UU.	6	0 “ningún día” a 6 “todos los días”	Unidimensional	220(78.19%) nT-r: 83(NM)	M = 45.2, DE = 8.8, NM	Adultos con obesidad
EES-D (Doğan et al., 2011)	Turquía	14	NM	Unidimensional	420(68.81%)	M = 26.26, DE = 9.60, 17-72	NM
CCE (Garaulet et al., 2012)	España	10	1 “nunca” a 4 “siempre”	1) Desinhibición, 2) Tipo de comida y 3) Culpa	354(84%) nT-r: 122(NM)	M = 39, DE = 12, NM	Adultos con obesidad
EAE (Rojas & García-Méndez, 2017)	México	36	1 “totalmente en desacuerdo” a 6 “totalmente de acuerdo”	1) Emoción, 2) Familia, 3) Indiferencia, 4) Cultura y 5) Efecto del alimento	795(54.46%)	M = 34 años, DE = 14.4, 18-60	NM
PNEES (Sultson et al., 2017)	Estonia	19	0 “nunca” a 4 “muy frecuentemente”	1) IE negativa y 2) IE positiva	E1: 531(100%) E2: 60(100%)	E1: M = 29.57, DE = 9.5, 15-61 E2: M = 21.6, DE = 2.85	NM
SEES (Meule et al., 2018)	Austria	20	1 “como mucho menos de lo usual” a 5 “como más de lo usual”	1) Felicidad, 2) Tristeza, 3) Enojo y 4) Ansiedad	E1: 285(78.95%) E2: 790(82.91%) E3: 623(53.77%)	E1: M = 25.4, DE = 9, 17-67 E2: M = 24.7, DE = 6.79, 15-65 E3: M = 33.5, DE = 18.2, 14-86	NM
FEED (Cassoli et al., 2021)	Italia	23	0 “nunca”/o “sin deseo de comer” a 4 “siempre”/“una necesidad imperante de comer”	1) Depresión, 2) Enojo y 3) Ansiedad	MG: 94(66%) MC: 234(NM) nT-r: 53(75.5%)	MG: M = 37, DE = 13.2, 18-65 años MC: NM	NM

Nota. EES-A = Emotional Eating Scale; E1 = estudio 1; E2 = estudio 2; nT-r = muestra de test-retest; EMAQ = Emotional Appetite Questionnaire; N/A = no aplica; DK = no saber qué responder; EMAQ-PE = emociones positivas; EMAQ-PS = situaciones positivas; EMAQ-NE = emociones negativas; EMAQ-NS = situaciones negativas; GB = grupo con bajo peso; GN = grupo con normopeso; GS = grupo con sobrepeso; IMC = índice de masa corporal; EOQ = Emotional Overeating Questionnaire; EES-D = Emotional Eating Scale; NM = no menciona; CCE = Cuestionario de Comedor Emocional; EAE = Escala de Alimentación Emocional; PNEES = Positive-Negative Emotional Eating Scale; IE = ingesta emocional; SEES = Salzburg Emotional Eating Scale; FEED = Florence Emotional Eating Drive; MG = muestra general; MC = muestra clínica.

Tabla 2. Análisis de propiedades psicométricas de las escalas de IE

Instrumento (autor, año)	Validez de contenido		Validez estructural		Consistencia interna		Estabilidad temporal		Validez de constructo	
	CMeto.	CMet.	CMeto.	CMet.	CMeto.	CMet.	CMeto.	CMet.	CMeto.	CMet.
EES-A (Arnou et al., 1995)	(I)	(?)	(I)	ACP/Varimax y Promax (-)	(V)	Tot $\alpha = .87$ /Enojo $\alpha = .78$, Ansiedad $\alpha = .78$ y Depresión $\alpha = .72$ (+)	(I)	2 semanas/ $r = 0.79^*$ (?)	(I)	(?)
EMAQ (Geliebter y Aversa, 2003)	(I)	(?)	(I)	NM (-)	(V)	1)EMAQ-PE $\alpha = .75$ 2)EMAQ-PS $\alpha = .57$ 3)EMAQ-NE $\alpha = .78$ 4)EMAQ-NS $\alpha = .65$ (-)	(I)	1)EMAQ-PE $r = .95$ 2)EMAQ-PS $r = .71$ 3)EMAQ-NE $r = .89$ 4)EMAQ-NS $r = .90$ (-)	NA	NA
EOQ (Masheby y Grillo, 2006)	(I)	(?)	(D)	ACP (-)	(V)	$\alpha = .85$ (+)	(I)	ICC = .62-.73** (?)	(I)	(+)
EES-D (Doğan et al., 2011)	(D)	(?)	(A)	AFE, ACP (-)	(V)	$\alpha = .94$ (+)	NA	NA	(I)	(?)
CCE(Garaulet et al., 2012)	(I)	(?)	(A)	ACP/Varimax (-)	(V)	Tot $\alpha = .77$ /Desinhibición $\alpha = .77$, Tipo de comida $\alpha = .66$ y Culpa $\alpha = .61$ (-)	(I)	$r = .70^{***}$ ICC = .40** (?)	(A)	(?)
EAE (Rojas y García-Méndez, 2017)	(I)	(?)	(A)	AFE/Varimax (-)	(V)	Tot $\alpha = .94$ /Emoción $\alpha = .96$, Familia $\alpha = .77$, Indiferencia $\alpha = .69$, Cultura $\alpha = .80$, Efecto del alimento $\alpha = .48$ (-)	NA	NA	NA	NA
PNEES (Sultson et al., 2017)	(I)	(?)	(A)	AFE/Promax (-)	(V)	Tot $\alpha = .89$ /IE negativa $\alpha = .94$ e IE positiva $\alpha = .80$ (+)	NA	NA	(V)	(+)
SEES (Meule et al., 2018)	(I)	(?)	(V)	AFE, ACP y AFC, MLE/ E2: CFI=.93, y RMSEA=.05 E3: CFI=.92, y RMSEA=.07 Promax (+)	(V)	E1:Felicidad $\alpha = .80$, Tristeza $\alpha = .77$, Enojo $\alpha = .77$ y Ansiedad $\alpha = .71$; E2:Felicidad $\alpha = .82$, Tristeza $\alpha = .77$, Enojo $\alpha = .75$ y Ansiedad $\alpha = .73$; E3:Felicidad $\alpha = .87$, Tristeza $\alpha = .85$, Enojo $\alpha = .86$ y Ansiedad $\alpha = .82$ (+)	NA	NA	(A)	(+)
FEED (Cassioli et al., 2021)	(I)	(?)	(A)	AFE, ACP/ promax y varimax (-)	(I)	M1 $\alpha = .89$ y M2 $\alpha = .96$ (+)	(D)	ICC = .38 - .91 $r = .93$ (-)	(A)	(+)

Nota. CMeto. = calidad metodológica; CMet. = calidad métrica; EES-A = Emotional Eating Scale; V = muy buena; A = adecuada; D = dudosa; I = inadecuado; (+) = calificación positiva; (?) = puntuación indeterminada; (-) = calificación negativa; EMAQ = Emotional Appetite Questionnaire; EMAQ-PE = emociones positivas; EMAQ-PS = situaciones positivas; EMAQ-NE = emociones negativas; EMAQ-NS = situaciones negativas; IE = ingesta emocional, NA = no aplica; EOQ = Emotional Overeating Questionnaire; ACP = Análisis de componentes principales; ICC = coeficiente intraclass; AFE = análisis factorial exploratorio; CCE = Questionario de Comedor Emocional; PNEES = Positive-Negative Emotional Eating Scale; SEES = Salzburg Emotional Eating Scale; MLE = maximum likelihood; CFI = comparative fit index; RMSEA = root mean square error of approximation; Tot = total; E1 = estudio 1; E2 = estudio 2; E3 = estudio 3; FEED = Florence Emotional Eating Drive; M1 = muestra 1, M2 = muestra 2. **p < .001. ***p < .0001

ubicaron en la categoría “inadecuada”. Respecto a la calidad métrica, la totalidad de los autoreportes obtuvieron una calificación perteneciente a la categoría “indeterminada”.

Validez estructural

Las técnicas más empleadas en los estudios analizados fueron el análisis factorial exploratorio (AFE) y el análisis de componentes principales (ACP). La rotación de los ítems únicamente fue reportada en seis investigaciones; dos utilizaron Varimax (Garaulet et al., 2012; Rojas & García-Méndez, 2017), dos Promax (Meule et al., 2018; Sultson et al., 2017) y las dos restantes usaron ambos tipos de rotación (Arnou et al., 1995; Cassioli et al., 2021). La varianza total explicada en los instrumentos osciló entre el 42.6% - 63%, siendo la PNEES (Sultson et al., 2017) la que explicó el mayor porcentaje.

Los instrumentos que lograron una calificación metodológica perteneciente a las categorías “muy buena” o “adecuada” fueron la EES-D (Doğan et al., 2011), el CCE (Garaulet et al., 2012), la EAE (Rojas & García-Méndez, 2017), la PNEES (Sultson et al., 2017), la SEES (Meule et al., 2018) y el FEED (Cassioli et al., 2021). Respecto a la calidad métrica, el instrumento que obtuvo la mayor calificación fue la SEES (Meule et al., 2018) dado que fue el único en realizar AFE y análisis factorial confirmatorio (AFC) e informar índices de bondad de ajuste.

Consistencia interna

La fiabilidad fue analizada en los nueve instrumentos a través del coeficiente alfa de Cronbach, el cual osciló entre .77 - .94 en las escalas que reportaron la consistencia interna global, cabe señalar que en el caso del EMAQ (Geliebter & Aversa, 2003) y la SEES (Meule et al., 2018) únicamente fueron reportados los coeficientes de consistencia interna de cada factor.

En cuanto a la calidad metodológica de esta propiedad ocho de los nueve instrumentos obtuvieron una calificación “muy buena”, siendo el FEED

(Cassioli et al., 2021) el único autoreporte que puntuó como inadecuado debido a que no fue reportada la consistencia interna para sus dimensiones. En tanto a la calidad métrica, en seis de los instrumentos: el EOQ (Masheb & Grilo, 2006), la EES-A (Arnou et al., 1995), la PNEES (Sultson et al., 2017), la SEES (Meule et al., 2018) y el FEED (Cassioli et al., 2021); se cumplió con el criterio de una consistencia interna $>.70$. El resto incumplieron dicha pauta con valores inferiores al deseado.

Estabilidad temporal

Los instrumentos que analizaron la fiabilidad *test-retest* fueron la EES-A (Arnou et al., 1995), el EMAQ (Geliebter & Aversa, 2003), el EOQ (Masheb & Grilo, 2006), el CCE (Garaulet et al., 2012) y el FEED (Cassioli et al., 2021). El rango del tamaño muestral utilizado en la prueba *test-retest* de los artículos osciló entre 47 y 122 personas con un tiempo considerado para la segunda aplicación entre una y dos semanas. Por otra parte, en cuatro de los cinco instrumentos se calculó el estadístico *r* de Pearson, el cual se encontró en un rango de .70 - .95 (Arnou et al., 1995; Cassioli et al., 2021; Garaulet et al., 2012; Geliebter & Aversa, 2003), en tanto que sólo en el EOQ (Masheb & Grilo, 2006) y el FEED (Cassioli et al., 2021) se proporcionó el coeficiente de correlación intraclase ICC = .62 - .73, $p < .01$; ICC = .38 - .91, respectivamente. Únicamente en el CCE (Garaulet et al., 2012) se mencionó haber obtenido un índice Kappa de .40 ($p < .0001$)

La mayoría de los estudios que analizaron la estabilidad temporal de los instrumentos obtuvieron una calidad metodológica insuficiente debido a que no se calculó el coeficiente intraclase o el índice Kappa ponderada así como una limitada información de la segunda aplicación, únicamente el FEED (Cassioli et al., 2021) alcanzó una calificación dudosa ya que, si bien proporciona el coeficiente intraclase, el tiempo entre aplicación fue apropiado y se proporciona el coeficiente de correlación no se menciona si los participantes estuvieron estables en el período intermedio así como tampoco las similitudes de las condiciones en las que se realizaron las evaluaciones.

Pruebas de hipótesis para la validez de constructo

En este análisis se incluye la validez convergente y divergente. La calidad metodológica más alta fue identificada en la PNEES (Sultson et al., 2017). En tanto que, cuatro de los nueve instrumentos obtuvieron una calificación positiva en su calidad métrica; el EOQ (Masheb & Grilo, 2006), la PNEES (Sultson et al., 2017), la SEES (Meule et al., 2018) y el FEED (Cassioli et al., 2021).

Otras propiedades de medida (Validez transcultural/medida de invarianza e interpretabilidad)

El SEES (Meule et al., 2018) fue el único autoreporte que reportó un análisis de invarianza en función del sexo. Respecto a la interpretabilidad, el CCE (Garaulet et al., 2012) es el único autoreporte que establece puntos de corte, siendo estos: 0-5 “comedor no emocional”, 6-10 “bajo comedor emocional”, 11-20 “comedor emocional” y 21-30 “comedor muy emocional”.

Discusión

El objetivo de la presente investigación fue realizar una revisión sistemática de la calidad psicométrica de los autoreportes de IE en población adulta. Para ello se analizaron nueve instrumentos. Se identificaron diferencias conceptuales respecto a la IE, lo cual puede deberse a la falta de un antecedente teórico específico para la conducta de IE; en el caso de la EES-A (Arnow et al., 1995), el EMAQ (Geliebter & Aversa, 2003), el EOQ (Masheb & Grilo, 2006), la EAE (Rojas & García-Méndez, 2017) y la EES-D (Doğan et al., 2011) señalaron como antecedentes explicativos de la IE la teoría de la restricción (Herman & Polivy, 1980) y la teoría psicosomática de la obesidad (Kaplan & Kaplan, 1957). Sin embargo, se ha referido que la IE puede presentarse de manera ocasional o habitual no necesariamente en casos de atracón alimentario (Macht & Simons, 2011). En el resto de instrumentos no se menciona la teoría en la cual se basaron para su desarrollo. Por otra parte, los instrumentos analizados muestran heterogeneidad en

torno a las dimensiones que los constituyen, así como la información que se puede obtener de los mismos.

La construcción de un instrumento es una etapa crucial en el proceso de elaboración del mismo. El contenido de los reactivos debe sustentarse en los principios de representatividad, relevancia, diversidad, claridad, sencillez y comprensibilidad (Muñiz et al., 2005). Llama la atención que en la mayoría de los instrumentos analizados no se informa la teoría en la que se basó la construcción del instrumento, cómo fueron construidos los reactivos y cómo se garantizó la validez de contenido.

La validez de contenido puede ser satisfactoria para distintas definiciones dadas a un mismo concepto, a la luz de la población y conceptualización que se especifique (Escobar-Pérez & Cuervo-Martínez, 2008); no obstante, debe considerarse que las diferencias conceptuales pueden generar limitaciones importantes cuando se pretende comparar resultados provenientes de instrumentos diferentes. Todos los instrumentos analizados miden la IE, no obstante, el análisis de los mismos mostró que su contenido es diverso, por lo que el significado de las puntuaciones es diferente. Por ello, es importante que al elegir uno de ellos se tenga un conocimiento de la definición conceptual y operacional del constructo.

En la EES-A (Arnow et al., 1995), la EES-D (Doğan et al., 2011), la EAE (Rojas & García-Méndez, 2017) y el FEED (Cassioli et al., 2021) se informaron elementos para el análisis de la validez relacionada con el contenido sin embargo, en ninguno se describió de forma detallada y precisa el proceso de construcción en el cual se involucra desde el diseño hasta el pilotaje del mismo, considerando: 1) la relevancia del constructo en una población y contexto específicos, 2) la exhaustividad de los aspectos claves del constructo y 3) la comprensibilidad de los reactivos; aspectos indagados con los profesionales y en la población para quien está dirigida la escala (Terwee et al., 2018). Por lo que, la totalidad de artículos obtuvo una calidad métrica “indeterminada”.

Respecto a la validez estructural se identificó que, en tres de los nueve instrumentos, EES-A (Arnow et al., 1995), EOQ (Masheb & Grilo, 2006) y CCE (Garaulet et al., 2012) se utilizó el ACP para identificar las dimensiones del constructo medido. Durante décadas

el uso de este análisis fue una práctica recurrente por tratarse aparentemente de un método sencillo y eficaz con supuesta equivalencia con el AFE (Lloret-Segura et al., 2014). Sin embargo, las recomendaciones actuales dan cuenta de la importancia de hacer uso del ACP o el AFE según corresponda al tipo de datos (Lloret-Segura et al., 2014).

Las investigaciones en las que se realizó análisis factorial fue la EES-D (Doğan et al., 2011), la EAE (Rojas & García-Méndez, 2017), la PNEES (Sultson et al., 2017), la SEES (Meule et al., 2018) y el FEED (Cassioli et al., 2021). Únicamente en el SEES (Meule et al., 2018) se realizó tanto el AFE como el AFC. Esto coincide con la recomendación de aplicar ambos análisis con la finalidad de explorar la estructura factorial y posteriormente confirmarla (Anderson & Gerbing, 1988) *surprisingly little research has examined the roles of a CEO's psychological traits. This study specifically examines the effects of CEOs' regulatory focus (promotion versus prevention.* En cuanto al tipo de rotación de los ítems se ha sugerido el uso de promax/oblicua en lugar de Varimax/ortogonal (Lloret-Segura et al., 2014).

En cuanto a la consistencia interna se encontró que para la totalidad de los instrumentos se calculó el coeficiente alfa de Cronbach. Es importante tener en cuenta que las recomendaciones psicométricas actuales señalan que el análisis de esta propiedad debe realizarse considerando el modelo de medida y los criterios para utilizar los coeficientes de consistencia interna (alfa u omega) con la finalidad de lograr estimaciones de fiabilidad más precisas (Viladrich et al., 2017). El alfa de Cronbach se considera un buen estimador cuando se trata de modelos congénicos con cargas factoriales altas y un gran número de ítems (Gu et al., 2013) mientras el coeficiente omega (jerárquico y total) son más adecuados cuando el modelo de medida implica efectos de método, factores espurios o errores correlacionados (Green & Yang, 2015). El proceso de selección de consistencia interna puede realizarse a través de tres etapas: 1) descripción de los ítems, 2) ajuste del modelo de medición y 3) elección del coeficiente de confiabilidad (Doval et al., 2023).

Respecto a la estabilidad temporal, en la EES-A (Arnold et al., 1995), el EMAQ (Geliebter & Aversa, 2003), el EOQ (Masheb & Grilo, 2006), el CCE (Garaulet et al., 2012) y el FEED (Cassioli et al., 2021) se reaplicó el

instrumento para evaluar su estabilidad temporal. De acuerdo con los parámetros establecidos por COSMIN (Prinsen et al., 2018) es preciso reportar la correlación intraclase o el peso Kappa ($\geq .70$) además proporcionar información sobre la estabilidad de los participantes en el período intermedio de la evaluación con respecto al constructo medido y precisar si las condiciones de las mediciones fueron similares. En las investigaciones analizadas el EOQ (Masheb & Grilo, 2006), el CCE (Garaulet et al., 2012) y el FEED (Cassioli et al., 2021) informaron la correlación intraclase y en ningún caso se proporcionó información sobre qué paso con los participantes en el periodo entre aplicaciones.

Respecto a la validez de constructo se identificaron deficiencias de acuerdo con los criterios COSMIN (Prinsen et al., 2018) debido a la falta de información sobre las propiedades psicométricas de los instrumentos con los que se analizó la relación con la escala de IE. Por ejemplo, no se especificó si los instrumentos utilizados consideraron muestras similares a las del autoreporte en cuestión. En los casos en los que se puntuó como dudosa la calidad métrica de esta propiedad se debió debido a que la hipótesis correlacional (convergente o discriminante) no fue explícitamente enunciada.

En tanto a la validez de criterio, de acuerdo con los criterios COSMIN en su última versión (Prinsen et al., 2018) el único estándar de oro para evaluar esta propiedad es la comparación de dos versiones de una misma escala (extendida y reducida). En los estudios analizados no se identificó el cumplimiento con tal criterio. Se considera oportuno dar cuenta de este análisis cuando se cuente con ambas versiones de los instrumentos, asimismo revisar la pertinencia de considerar otros procedimientos que permiten generar una evidencia sobre esta propiedad métrica.

En la presente investigación se encontró que las escalas que obtuvieron las mayores puntuaciones respecto a su calidad metodológica y su calidad métrica fueron la PNEES (Sultson et al., 2017) y la SEES (Meule et al., 2018). Cabe señalar que son instrumentos creados recientemente, por lo que es de esperarse que incluyan análisis estadísticos desarrollados en los últimos años, por tanto, más apegados a las consideraciones métricas y metodológicas actuales.

Las limitaciones de la presente revisión radican en que la información analizada se reduce exclusivamente

a la presentada en los estudios originales. Para los fines de la presente revisión no se consideraron todas las evidencias que pudieran haber de los instrumentos analizados, por lo que se sugiere que a partir de la presente se realicen revisiones sistemáticas específicas para cada instrumento incluyendo todos los estudios en los que se hayan analizado sus propiedades métricas en diferentes poblaciones. La PNEES (Sultson et al., 2017) y la SEES (Meule et al., 2018) obtuvieron las mejores puntuaciones de acuerdo con los estándares y criterios COSMIN, por lo que es necesario continuar generando evidencia sobre su calidad métrica. Otra limitante del presente estudio es referente al idioma, dado que fueron excluidos los artículos que no estuvieran escritos en inglés o español.

La presente investigación constituye el primer análisis de los instrumentos que miden la IE considerando los criterios COSMIN (Prinsen et al., 2018) para evaluar la calidad metodológica y métrica de los autoreportes. El análisis realizado permite concluir que: 1) es importante conocer la teoría retomada para el desarrollo de los instrumentos de IE ya que existen diferencias conceptuales entre los autoreportes; la mayoría de ellos son dimensionales y con una mayor frecuencia evalúan emociones específicas (e. j., Enojo, Ansiedad, Depresión, Felicidad). Asimismo, los instrumentos muestran discrepancias en las opciones de respuesta, siendo las opciones más frecuentes las que indagan por: la modificación en la conducta de comer (e. j., 1 “como mucho menos de lo usual” a 5 “como mucho más de lo usual”), la ocurrencia con la que se ve alterada la conducta alimentaria (e. j., 0 “nunca” a 4 “muy frecuentemente”) y la modificación en el apetito (1 “no deseo comer” a 5 “una necesidad imperante de comer”). Por lo que la información obtenida depende del instrumento utilizado; 2) es preciso aportar evidencia de la validez relacionada con el contenido de forma detallada, particularmente de aquellos en los que no se proporcionó información al respecto en sus versiones originales; 3) la PNEES (Sultson et al., 2017) y la SEES (Meule et al., 2018) lograron las mejores puntuaciones tanto en la evaluación metodológica como métrica y 4) es importante considerar que la evaluación de las propiedades métricas de un instrumento debe ser continua por lo que posteriores estudios deben atender las áreas de oportunidad encontradas en este análisis.

Referencias

*Artículos incluidos en la revisión sistemática

- Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin*, 103(3), 411–423. <https://doi.org/10.1037/0033-2909.103.3.411>
- *Arnow, B., Kenardy, J., & Agras, W. S. (1995). The Emotional Eating Scale: the development of a measure to assess coping with negative affect by eating. *International Journal of Eating Disorders*, 18(1), 79–90. doi: 10.1002/1098-108x(199507)18:1<79::aid-eat2260180109>3.0.co;2-v.
- Bongers, P., & Jansen, A. (2016). Emotional eating is not what you think it is and emotional eating scales do not measure what you think they measure. *Frontiers in Psychology*, 7(1932). <https://doi.org/10.3389/fpsyg.2016.01932>
- Braden, A., Musher-eizenman, D., Watford, T., Emley, E., & Musher-eizenman, D. (2018). Eating when depressed, anxious, bored, or happy: Are emotional eating types associated with unique psychological and physical health correlates? *Appetite*, 125, 410–417. <https://doi.org/10.1016/j.appet.2018.02.022>
- *Cassioli, E., Calderani, E., Fioravanti, G., Lazzeretti, L., Rotella, C. M., Rossi, E.,... Rotella, F. (2021). The Florence Emotional Eating Drive (FEED): A validation study of a self-report questionnaire for emotional eating. *Eating and Weight Disorders*. <https://doi.org/10.1007/s40519-021-01216-2>
- *Doğan, T., Tekin, E. G., & Katrancıoğlu, A. (2011). Feeding your feelings: A self-report measure of emotional eating. *Procedia - Social and Behavioral Sciences*, 15, 2074–2077. <https://doi.org/10.1016/j.sbspro.2011.04.056>
- Doval, E., Viladrich, C., & Angulo-Brunet, A. (2023). Coefficient alpha: The resistance of a classic. *Psicothema*, 35(1), 5–20. <https://doi.org/10.7334/psicothema2022.321>
- Escobar-Pérez, J., & Cuervo-Martínez, Á. (2008). Validez de contenido y juicio de expertos: Una Aproximación a su utilización. *Avances en Medición*, 6, 27–36. https://www.researchgate.net/publication/302438451_Validez_de_contenido_y_juicio_de_expertos_Una_aproximacion_a_su_utilizacion
- Franco Paredes, K., Mancilla Díaz, J. M., Vázquez Arévalo, R., Álvarez Rayón, G., & López Aguilar, X. (2011). El papel del perfeccionismo en la insatisfacción corporal, la influencia sociocultural del modelo de delgadez y los síntomas de trastorno del comportamiento alimentario. *Universitas Psychologica*, 10(3), 829840. <http://www.redalyc.org/html/647/64722377015/>

- Frayn, M., & Knäuper, B. (2017). Emotional Eating and Weight in Adults: a Review. *Current Psychology*, 37(4), 924–933. <https://doi.org/10.1007/s12144-017-9577-9>
- Ganley, R. M. (1988). Emotional eating and how it relates to dietary restraint, disinhibition, and perceived hunger. *International Journal of Eating Disorders*, 7(5), 635–647. [https://doi.org/10.1002/1098-108X\(198809\)7:5<635::AID-EAT2260070507>3.0.CO;2-K](https://doi.org/10.1002/1098-108X(198809)7:5<635::AID-EAT2260070507>3.0.CO;2-K)
- *Garaulet, M., Canteras, M., Morales, E., López-Guimera, G., Sánchez-Carracedo, D., & Corbalán-Tatua, M. D. (2012). Validation of a questionnaire on emotional eating for use in case of obesity; the Emotional Eater Questionnaire (EEQ). *Nutrición Hospitalaria*, 27(2), 645–651. <https://doi.org/10.3305/nh.2012.27.2.5659>
- *Geliebter, A., & Aversa, A. (2003). Emotional eating in overweight, normal weight, and underweight individuals. *Eating Behaviors*, 3(4), 341–347. doi: 10.1016/S1471-0153(02)00100-9
- Green, S. B., & Yang, Y. (2015). Evaluation of dimensionality in the assessment of internal consistency reliability: Coefficient alpha and omega coefficients. *Educational Measurement: Issues and Practice*, 34(4), 14–20. <https://doi.org/10.1111/EMIP.12100>
- Gu, F., Little, T. D., & Kingston, N. M. (2013). Misestimation of reliability using coefficient alpha and structural equation modeling when assumptions of tau-equivalence and uncorrelated errors are violated. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 9(1), 30–40. <https://doi.org/10.1027/1614-2241/a000052>
- Herman, C. P., & Polivy, J. (1980). Restrained eating. In A. J. Stunkard (Ed.), *Obesity* (pp. 208–225). Saunders.
- Kaplan, H. I., & Kaplan, H. S. (1957). The psychosomatic concept of obesity. *The Journal of Nervous and Mental Disease*, 125(2), 181–201. doi:10.1097/00005053-195704000-00004
- Koball, A. M., Meers, M. R., Storfes-Isser, A., Domoff, S. E., & Musher-Eizenman, D. R. (2012). Eating when bored: Revision of the emotional eating scale with a focus on boredom. *Health Psychology*, 31(4), 521–524. <https://doi.org/10.1037/a0025893>
- Lane, S., Raymond, M.R., & Haladyna, T. M. (2016). *Handbook of test development* (2nd ed.). New York: Routledge.
- Lloret-Segura, S., Ferreres-Traver, A., Hernández-Baeza, A., & Tomás-Marco, I. (2014). El análisis factorial exploratorio de los ítems: Una guía práctica, revisada y actualizada. *Anales de Psicología*, 30(3), 1151–1169. <http://dx.doi.org/10.6018/analesps.30.3.199361>
- Macht, M., & Simons, G. (2011). Emotional eating. In I. Nyklíček, Ad. Vingerhoets y M. Zeelenberg (Eds.), *Emotion Regulation and Well-Being* (pp. 281–295). Springer. <https://doi.org/10.1007/978-1-4419-6953-8>
- *Masheb, R. M., & Grilo, C. M. (2006). Emotional overeating and its associations with eating disorder psychopathology among overweight patients with binge eating disorder. *International Journal of Eating Disorders*, 39(2), 141–146. <https://doi.org/10.1002/eat.20221>
- *Meule, A., Reichenberger, J., & Blechert, J. (2018). Development and preliminary validation of the Salzburg Emotional Eating Scale. *Frontiers in Psychology*, 9(88), 1–10. <https://doi.org/10.3389/fpsyg.2018.00088>
- Mokkink, L. B., de Vet, H. C. W., Prinsen, C. A. C., Patrick, D. L., Alonso, J., Bouter, L. M., & Terwee, C. B. (2018). COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. *Quality of Life Research*, 27(5), 1171–1179. <https://doi.org/10.1007/s11136-017-1765-4>
- Mokkink, L. B., Prinsen, C. A. C., Bouter, L. M., de Vet, H. C. W., & Terwee, C. B. (2016). The consensus-based standards for the selection of health measurement instruments (COSMIN) and how to select an outcome measurement instrument. *Brazilian Journal of Physical Therapy*, 20(2), 105–113. <https://doi.org/10.1590/bjpt-rbf.2014.0143>
- Muñiz, J., Fidalgo, A. M., García-Cueto, E., Martínez, R., & Moreno, R. (2005). *Análisis de los ítems*. España: La Muralla.
- Uzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan—a web and mobile app for systematic reviews. *Systematic Reviews*, 5(210), 1–10. <https://doi.org/10.1186/s13643-016-0384-4>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *The British Medical Journal*, 372(71), 1–9. <https://doi.org/10.1136/BMJ.N71>
- Prinsen, C. A. C., Mokkink, L. B., Bouter, L. M., Alonso, J., Patrick, D. L., de Vet, H. C. W., & Terwee, C. B. (2018). COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of Life Research*, 27, 1147–1157. <https://doi.org/10.1007/s11136-018-1798-3>
- Reichenberger, J., Schnepfer, R., Arend, A.-K., & Blechert, J. (2020). Emotional eating in healthy individuals and patients with an eating disorder: evidence from psychometric, experimental and naturalistic studies. *Proceedings of the Nutrition Society*, 79, 290–299. <https://doi.org/10.1017/S0029665120007004>
- *Rojas, A. T., & García-Méndez, M. (2017). Construcción de una Escala de Alimentación Emocional. *Revista Iberoamericana de Diagnóstico y Evaluación*, 3(45), 85–95. <https://doi.org/10.21865/RIDEP45.3.07>
- Smith, J., Ang, X. Q., Giles, E. L., & Traviss-Turner, G. (2023). Emotional Eating Interventions for Adults Living

- with Overweight or Obesity: A Systematic Review and Meta-Analysis. *International Journal of Environmental Research and Public Health*, 20(3). <https://doi.org/10.3390/ijerph20032722>
- *Sultson, H., Kukk, K., & Akkermann, K. (2017). Positive and negative emotional eating have different associations with overeating and binge eating: Construction and validation of the Positive-Negative Emotional Eating Scale. *Appetite*, 116, 423–430. <https://doi.org/10.1016/j.appet.2017.05.035>
- Sze, K. Y. P., Lee, E. K. P., Chan, R. H. W., & Kim, J. H. (2021). Prevalence of negative emotional eating and its associated psychosocial factors among urban Chinese undergraduates in Hong Kong: a cross-sectional study. *BMC Public Health*, 21(1), 1–10. <https://doi.org/10.1186/s12889-021-10531-3>
- Terwee, C. B., Prinsen, C. A. C., Chiarotto, A., De Vet, H. C. W., Westerman, M. J., Patrick, D. L., Mookink, L. B. (2018). COSMIN standards and criteria for evaluating the content validity of health-related patient-reported outcome measures: A delphi study. *Quality of Life Research*, 27(5), 1159–1170. doi: 10.1007/s11136-018-1829-0
- van Strien, T., Frijters, J. E. R., Bergers, G. P. A., & Defares, P. B. (1986). The Dutch Eating Behavior Questionnaire (DEBQ) for assessment of restrained, emotional, and external eating behavior. *International Journal of Eating Disorders*, 5(2), 295–315. [https://doi.org/10.1002/1098-108X\(198602\)5:2<295::AID-EAT2260050209>3.O.CO;2-T](https://doi.org/10.1002/1098-108X(198602)5:2<295::AID-EAT2260050209>3.O.CO;2-T)
- Vázquez-Vázquez, V. M., Bosques-Brugada, L. E., Guzmán-Saldaña, R. M. E., Romero-Palencia, A., Reyes-Jarquín, K., & Franco-Paredes, K. (2019). Revisión del constructo y fundamentos teóricos de la alimentación emocional. *Educación y Salud Boletín Científico de Ciencias de La Salud Del ICSa*, 8(15), 255–26.
- Viladrich, C., Angulo-Brunet, A., & Doval, E. (2017). Un viaje alrededor de alfa y omega para estimar la fiabilidad de consistencia interna. *Anales de Psicología*, 33(3), 755–782. <https://doi.org/10.6018/analesps.33.3.268401>