

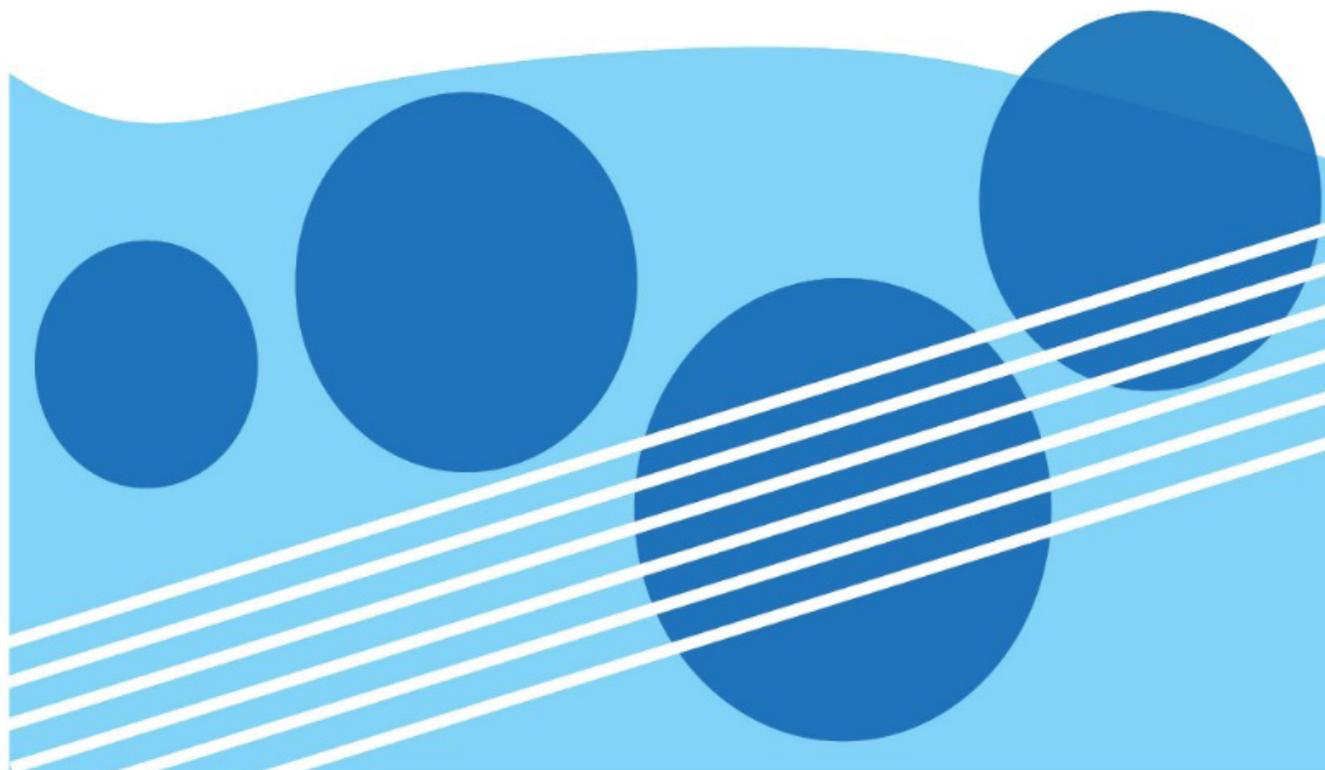


Revista Evaluar

Laboratorio de Educación Psicológica y Educativa
Facultad de Psicología de la Universidad Nacional de Córdoba

2019

VOL 19- N°2
ISSN1667-4545



Escala Argentina de Dificultades Alimentarias en Niños (EADAN): Propiedades psicométricas

Argentine Scale of Feeding Difficulties in Children (EADAN): Psychometric properties

Maria Paulina Hauser * ¹, Ruth Alejandra Taborda ¹,
Alicia Oiberman ², Maria Ramsay ³

1 - Facultad de Psicología, Universidad Nacional de San Luis.

2 - CIIPME - CONICET. Facultad de Psicología, Universidad de Buenos Aires.

3 - Departamento de Psicología, Hospital de Niños de Montreal. Departamento de Pediatría,
Universidad McGill, Montreal, Quebec.

Recibido: 31/01/2019 Revisado: 03/03/2019 Aceptado: 06/03/2019

Resumen

El objetivo de esta investigación fue adaptar y validar en Argentina un cuestionario para padres que detecta dificultades en la alimentación, las cuales son muy comunes entre los niños pequeños. Se analizaron las propiedades psicométricas de la Escala Argentina de Dificultades Alimentarias en Niños (EADAN) en una muestra argentina de 263 niños de entre los 6 meses y los 6 años 11 meses de edad. Existen diferencias significativas entre la media por ítem y la media del puntaje total en los grupos con dificultades y sin dificultades en la alimentación. La confiabilidad de la escala es buena, con un valor $\alpha = .79$. Se realizó un análisis factorial exploratorio que arrojó dos factores que representan el 49.4% de la varianza. Se compararon los puntajes obtenidos por niños prematuros y nacidos a término. Se concluye acerca de la validez del instrumento para la detección de dificultades en la alimentación en niños argentinos.

Palabras clave: *problemas de alimentación, niños, cuestionario, Argentina, niños prematuros, niños nacidos a término*

Summary

The objective of this study was to adapt and validate in Argentina a questionnaire for parents that detects feeding difficulties, which are very common among young children. Psychometric properties of the Argentine Scale of Feeding Difficulties in Children (EADAN) were analyzed in an Argentinian sample of 263 children, aged 6 months to 6 years 11 months. There are significant differences between the item mean and the total score mean between the groups with feeding difficulties and without feeding difficulties. The reliability of the scale is good, with an $\alpha = .79$. An exploratory factor analysis showed two factors that represent 49.4% of the variance. Scores obtained by premature and full-term infants were compared. It is concluded that this instrument is valid for the detection of feeding difficulties in Argentine children.

Key words: *feeding problems, children, questionnaire, Argentina, premature, full-term*

Introducción
Metodología
Resultados
Discusión
Referencias

*Correspondencia a: María Paulina Hauser. Facultad de Psicología, Universidad Nacional de San Luis. CIIPME - CONICET. Tte. Gral. Juan Domingo Perón 2158, C1040AAH CABA. Buenos Aires, Argentina. Tel: +54 (011) 4953-1477/3541. E-mail: mphauser@unsl.edu.ar

Nota de autor: Se agradece a las instituciones en las que se realizó la investigación y al Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) por el financiamiento de la misma mediante una Beca Postdoctoral.

Cómo citar este artículo: Hauser, M. P., Taborda, R. H., Oiberman, A., & Ramsay, M. (2019). Escala Argentina de Dificultades Alimentarias en Niños (EADAN): Propiedades psicométricas. *Revista Evaluar*, 19(2), 1-14. Recuperado de <https://revistas.unc.edu.ar/index.php/revaluar>

Introducción

Los problemas de alimentación son muy comunes entre los niños pequeños, tienen un impacto negativo en su desarrollo y están asociados a un mayor estrés durante el momento de la alimentación, pobre crecimiento y retraso en el desarrollo cognitivo (van Dijk, Timmerman, Martel, & Ramsay, 2011). Las dificultades en la alimentación y la ganancia de peso en la infancia generan intensas preocupaciones en los padres, por lo que son motivos frecuentes de consulta en los servicios pediátricos. Entre los niños con un desarrollo típico, se estima que las dificultades de alimentación se presentan en el 25% al 45% de los niños con desarrollo típico, mientras que se dan en un 30% al 80% en niños con trastornos del desarrollo (Arvedson, como se citó en Barton, Bickell, & Fucile, 2017).

Las dificultades en torno a la alimentación que pueden presentar los niños abarcan desde la ingesta en cantidades limitadas o sólo de cierto tipo de alimentos hasta la imposibilidad de comer por sí mismos y de manejar la comida o los utensilios apropiadamente. Si bien en muchos casos estas complicaciones pasan desapercibidas, pueden afectar de manera significativa la salud física y psicosocial de los niños y sus padres (Adamson, Morawska, & Sanders, 2013). En un estudio realizado en Singapur se entrevistó a 407 cuidadores, quienes informaron que el comportamiento quisquilloso al comer y las dificultades aparecieron alrededor del primer año de vida de los niños. La mayor tendencia a presentar dificultades en la alimentación se observó en niños de entre 3 y 10 años, en niños con padres profesionales y en aquellos que tenían antecedentes familiares de dificultades en la alimentación. El comportamiento quisquilloso al comer y las dificultades en la alimentación se asociaron significativamente con el estrés de los cuidadores durante las comidas y

el impacto negativo en las relaciones familiares (Goh & Jacob, 2012).

Los problemas de alimentación en la infancia se explican desde un modelo biopsicosocial, lo que supone que son el resultado del interjuego entre factores fisiológicos y psicosociales (Johnson & Harris, 2004; Rommel, De Meyer, Feenstra, & Veereman-Wauters, 2003). Los factores fisiológicos que pueden contribuir al desarrollo de problemas en la alimentación incluyen la prematuridad, el poco apetito y las dificultades en la sensomotricidad oral, mientras que los factores psicosociales incluyen el temperamento del niño y las dificultades en la relación madre-hijo (van Dijk et al., 2011). La relación que se establece entre la madre y el niño en torno a la alimentación está influenciada tanto por factores fisiológicos como interactivos (Ramsay, 2013).

El factor conductual es señalado como central en la aparición de dificultades en la alimentación de los niños junto con ciertos factores ambientales y conductuales de los padres que parecen predecir la mayor o menor aparición de dificultades. Entre las conductas parentales que aumentarían las dificultades se señalan: proporcionar comidas alternativas, amenazar y presionar al niño para comer, el castigo físico y las reprimendas (Adamson et al., 2013). Las estrategias o conductas utilizadas usualmente por los padres para facilitar o limitar la ingesta de ciertos alimentos se denominan prácticas de alimentación, dentro de las cuales las más estudiadas son la restricción y la presión para comer. Por lo general, estas prácticas responden a características o conductas específicas del niño, como por ejemplo ser quisquilloso para comer (Blisseth, 2011). El estilo adoptado para alimentar a los niños parece estar influido en cierta medida por el nivel educativo materno. En una investigación realizada en el Reino Unido, se obtuvo que las madres con mayor nivel de estudios suelen tener más control sobre la comida y

recurren significativamente menos a la *alimentación emocional*, en la cual los alimentos se ofrecen para calmar al niño en lugar de saciar su apetito. Las estrategias de alimentación emocional y de alimentación instrumental, en la que la comida es ofrecida como recompensa o retirada a modo de castigo, suelen ser perjudiciales para el manejo adecuado del peso de los niños (Saxton, Carnell, van Jaarsveld, & Wardle, 2009).

En una investigación reciente realizada en Dinamarca, se investigaron los factores perinatales de riesgo asociados a la aparición de desórdenes en la alimentación en niños de 0 a 3 años. Los investigadores hallaron que los niños nacidos antes de las 28 semanas de edad gestacional (EG), los pequeños para su edad gestacional y los infantes con malformaciones tenían el mayor riesgo de presentar este tipo de desórdenes. Además, la pertenencia al sexo femenino, el hábito materno de fumar durante el embarazo y el estatus de inmigrante y de primer hijo parecerían estar asociados con el incremento del riesgo de desórdenes en la alimentación en la población estudiada. En cuanto a la prevalencia de los desórdenes en la alimentación, se encontró que la mayoría de los niños eran diagnosticados durante los primeros 12 meses de vida, en particular en los primeros 4 meses. Los prematuros fueron quienes tenían más probabilidades de ser diagnosticados, probablemente debido a problemas neurológicos, como consecuencia de la alimentación mediante sonda naso u orogástrica y de las características de la relación temprana madre-hijo (Hvelplund, Hansen, Vinkel-Koch, Andersson, & Skovgaard, 2016). En una investigación reciente conducida en niños prematuros (Van Dijk, Bruinsma, & Hauser, 2016), hallaron que a medida que los padres informaban mayores problemas en la alimentación, los niños observados rechazaban en mayor medida los alimentos y comían de manera menos eficiente e independiente. Los niños prematuros de este estudio presenta-

ron puntajes que indicaban ausencia de problemas en la alimentación o problemas leves.

Dado que la comunidad científica no posee aún una definición aceptada de manera universal respecto de las dificultades en la alimentación, los investigadores han utilizado diversos términos para referirse a los comportamientos que indicarían dichos problemas (Van Dijk et al., 2011). Se señalan como indicadores de dificultades en la alimentación en niños: el rechazo de la comida (Lindberg, Bohlin, & Hagekull, 1991), la irregularidad al comer (McDermott et al., 2008), la desobediencia durante la alimentación (Sanders, Patel, Le Grice, & Shepherd, 1993) y la negatividad a la hora de comer (Johnson & Harris, 2004). Empujar los alimentos fuera de la boca, tener arcadas, guardar comida en la boca sin tragarla y llorar también han sido descritos como determinantes para considerar si un niño tiene problemas de alimentación (Chung et al., como se citó en Fernández-Gallardo, Rojas-Contreras, & Vargas-Keith, 2017). Si bien los comportamientos problemáticos en torno a la alimentación (por ejemplo, girar la cabeza, arquear el cuerpo o llorar) tienden a ser reacciones a estímulos internos en un primer momento (falta de hambre, falta de capacidad de succión), también pueden llegar a ser condicionados por la asociación a las señales externas y sociales (por ejemplo, engaños de parte de los padres). La detección temprana puede mejorar los comportamientos durante la alimentación y las interacciones a la hora de la comida, que a su vez ayudan a promover la independencia y otras habilidades en el niño (Ramsay, 2013).

Debido a la prevalencia de los problemas de alimentación y sus consecuencias negativas en el desarrollo de los niños pequeños, resulta de gran importancia que los profesionales que trabajan en primera infancia cuenten con instrumentos que permitan la rápida detección de dificultades en la alimentación en niños. Con el propósito de res-

ponder a esta vacancia, Ramsay, Martel, Porporino y Zygmuntowicz (2011) diseñaron en Canadá la escala Montreal Children's Hospital Feeding Scale (MCH-FS). Este es un instrumento de fácil administración, cuyo propósito es detectar rápidamente dificultades en la alimentación en niños pequeños en una consulta breve.

La MCH-FS está compuesta por 14 ítems que fueron contruidos por psicólogos expertos en el trabajo con niños con problemas en la alimentación. Estos ítems indagan en diferentes aspectos de la alimentación: lo oromotor, lo sensoriomotor, el apetito, las preocupaciones maternas sobre la alimentación, los comportamientos en el momento de la alimentación, las estrategias utilizadas por la madre y las reacciones familiares ante la alimentación del niño. Se le solicita a la persona encargada de alimentar al niño que puntúe los 14 ítems en una escala Likert de siete puntos, con puntos de atracción en cada extremo. El cuidador puntúa cada ítem de acuerdo a la frecuencia o el nivel de dificultad en un comportamiento particular o según el nivel de preocupación parental. Se obtiene un puntaje total a partir de la suma de todos los ítems. Ese puntaje bruto luego se convierte en un puntaje T que indicaría dificultades leves, moderadas o severas en la alimentación. A mayor puntaje, mayor la gravedad de las dificultades.

El instrumento fue sometido a estudios de validez y confiabilidad. A fin de establecer la validez de constructo, se administró el cuestionario a 198 madres de niños que asistían al consultorio pediátrico (muestra normativa) y a 174 madres de niños derivados a una clínica especializada en desórdenes de la alimentación (muestra clínica). Se obtuvo una excelente validez de constructo, ya que al comparar las puntuaciones medias de las muestras normativas y clínicas se observaron claras diferencias ($M = 32.65$, $DE = 12.73$ contra $M = 60.48$, $DE = 13.04$ respectivamente; $p < .01$).

La confiabilidad test-retest fue buena para ambos grupos (normativo $r = .84$ y clínico $r = .92$). El análisis de factor realizado mediante el análisis de componentes principales evidenció que un solo factor representaba al 48% de la varianza, lo que sugiere que el puntaje total obtenido en la escala puede ser utilizado como una medida de los problemas de alimentación. Los resultados evidencian que los niños en el grupo clínico presentan los mismos problemas en la alimentación que los niños del grupo control, pero en mayor frecuencia o de manera más severa (Ramsay et al., 2011).

Este instrumento MCH-FS también fue traducido y estandarizado en los Países Bajos y en Tailandia. La versión holandesa del instrumento, denominada Screeningslijst Eetgedrag Peuters (SEP; Van Dijk et al., 2011), fue sometida a estudios de validez y confiabilidad en una muestra normativa de 1448 niños entre los 6 meses y los 4 años que residían en la provincia de Groninga de los Países Bajos. Los resultados evidencian una buena confiabilidad (alfa de Cronbach $\alpha = .84$). En contraste con la investigación de los creadores de la escala, a partir del análisis factorial exploratorio mediante componentes principales se obtuvieron dos factores que se corresponderían significativamente con 12 de los 14 ítems de la MCH-FS y que representan una varianza del 43%. La correlación entre los dos factores es alta, de .69. Los autores consideran que el primer factor reúne los *comportamientos negativos durante las comidas* mientras que el segundo factor incluye aquellos ítems que indicarían *causas y consecuencias negativas de las dificultades en la alimentación*. A partir del análisis de las propiedades psicométricas, se puede concluir que la SEP es un instrumento válido para la detección de problemas en la alimentación en niños.

La versión tailandesa de la MCH-FS (Benjasuwantep, Rattamongkolgul, & Ramsay, 2015) fue desarrollada a partir de una muestra de 200

niños de entre los 12 y los 48 meses. La confiabilidad fue alta, con un alfa de Cronbach de .83. El análisis factorial se realizó utilizando el análisis de componentes principales, lo que resultó en tres factores que dan cuenta del 52.3% de la varianza, con una correlación moderada entre los tres factores. Los primeros dos factores están relacionados con *causas de comportamientos negativos durante las comidas y preocupaciones y reacciones parentales*, el tercer factor contempla *efectos de las situaciones de alimentación negativas*, con una correlación moderada entre los tres factores. Es interesante señalar que los autores debieron realizar una modificación en el ítem 5 introduciendo la opción *Nunca finaliza la comida*.

A partir de la revisión de los instrumentos de los que se dispone en nuestro medio, se tomó conocimiento de la ausencia de pruebas estandarizadas en nuestra población para la detección de las dificultades en la alimentación en la primera infancia. Es por esto que el objetivo de esta investigación fue adaptar y validar en Argentina el cuestionario para padres MCH-FS (Ramsay et al., 2011). El presente trabajo se propone analizar las propiedades psicométricas de la versión local denominada Escala Argentina de Dificultades Alimentarias en Niños (EADAN) en una muestra obtenida de la población argentina.

Metodología

Participantes

Para la selección de participantes se instrumentó un proceso bietápico. En la primera etapa se escogió en la ciudad de San Luis, provincia de San Luis, una clínica privada y la Maternidad Provincial Dra. Teresita Baigorria. En la segunda etapa se seleccionó una Maternidad pública en la ciudad de Avellaneda, provincia de Buenos Aires. Dichas instituciones fueron escogidas por ser cen-

tros de referencia en las zonas en las que se llevó a cabo la investigación.

La muestra está formada por 263 niños de entre los 6 meses y los 6 años y 11 meses de edad. Se incluyeron 177 niños nacidos a término y 86 niños prematuros. Los niños prematuros se clasificaron según su edad gestacional (EG) al nacer en prematuros extremos (menos de 28 semanas de EG al nacer), muy prematuros (entre 28 y 31 semanas de EG) y moderados (entre 32 y 36 semanas de EG). Se excluyeron niños que habían recibido previamente un diagnóstico de desorden en la alimentación o se encontraban en tratamiento debido a problemas de alimentación. Se solicitó el consentimiento informado a los padres de todos los niños involucrados.

Instrumentos

Se utilizó la Escala Argentina de Dificultades Alimentarias en Niños (EADAN), versión argentina del cuestionario MCH-FS (Ramsay et al., 2011). Para validar el instrumento, primero fue necesario adaptarlo culturalmente a la población local en la que se pretendía utilizar y luego volver a medir sus propiedades psicométricas. El MCH-FS fue traducido con el permiso de la autora creadora del instrumento, quien además es una de las autoras del presente artículo. La versión en inglés del MCH-FS fue traducida al español por un traductor bilingüe cuya lengua materna es el español. Luego fue retraducida al inglés por dos traductores bilingües, uno cuya lengua materna es el inglés y en el otro caso, el español. Estas versiones fueron comparadas entre sí para evaluar la concordancia con la versión original. Fue necesario introducir ciertas modificaciones en la versión en español a fin de utilizar palabras que fueran más frecuentes en el uso cotidiano y, por ende, más fáciles de comprender. Estos cambios fueron

consultados con la autora original, quien consideró que eran pertinentes si contribuían a mejorar la comprensión en el idioma español. La versión final en español fue enviada a tres expertos en la temática a fin de estudiar la pertinencia de los ítems indagados en nuestra población local. Según los expertos, todos los ítems reflejan nuestra cultura, así como las dificultades en torno a la alimentación, y se encuentran redactados de forma tal que pueden ser comprendidos por los padres.

Junto con el cuestionario EADAN, se les entregó a los padres de los niños un cuestionario sociodemográfico, elaborado por los autores de este artículo, que permitió obtener información sobre la edad del padre o madre, nivel educativo, situación laboral y lugar de residencia. También permitió recabar información sobre la edad del niño, la fecha de nacimiento, la edad gestacional al nacer, el peso y la talla de nacimiento, si estuvo internado en la UCIN (tiempo de internación, motivo, asistencia respiratoria utilizada, alimentación por sonda), peso y talla actual del niño, si recibe algún tipo de suplemento nutricional, si padece de alguna enfermedad crónica (medicación), si come junto con la familia o por separado, el nivel de desarrollo alcanzado y antecedentes familiares de trastornos de la alimentación.

Procedimiento

Los cuestionarios fueron entregados a los padres de los niños que asistían a instituciones de salud en las ciudades de San Luis y Avellaneda en el periodo comprendido entre junio de 2017 y noviembre de 2018. Las tres instituciones poseen características diferentes. Una es una institución privada ubicada en la zona céntrica de la ciudad de San Luis, a la que asisten en su gran mayoría personas que poseen obra social y tienen trabajos estables. Otra es una institución pública

de referencia a nivel provincial, por lo que asisten personas de diversos lugares del interior que, en general, no poseen obra social y se encuentran en una situación socioeconómica difícil. Esta institución recibe además a niños nacidos prematuros o de alto riesgo en su Consultorio de Seguimiento para Niños de Alto Riesgo. La tercera institución se encuentra ubicada en el primer cordón del conurbano bonaerense y asiste a una población que no posee, en su amplia mayoría, obra social y se encuentra en situación de vulnerabilidad social, económica y cultural. Este hospital también posee un Consultorio de Seguimiento al que asisten niños de Alto Riesgo.

Los padres podían responder el cuestionario en el tiempo de espera para ingresar a la consulta pediátrica, o bien luego de la misma. Si bien se trata de un cuestionario autoadministrado, en todos los casos el primer investigador asistió a los padres durante la respuesta del mismo a fin de aclarar si surgía alguna duda y asegurarse de que los cuestionarios fueran completados en su totalidad. Previo a la respuesta del cuestionario se les informó a los sujetos acerca del objetivo de la investigación, así como que su participación era voluntaria y anónima, y se les solicitó la firma del consentimiento informado. Todos los cuestionarios fueron ingresados de forma manual en el paquete estadístico SPSS versión 21 (IBM, 2012). En los ítems 1, 3, 4, 8, 10, 12 y 13 el puntaje se ingresó de manera invertida tal como indica la versión original para la puntuación; en el caso de estos ítems se utilizaron para el análisis los valores invertidos.

Análisis estadístico

El análisis estadístico de los ítems individuales y del puntaje total de la escala se realizó en SPSS versión 21 (IBM, 2012). El análisis fac-

torial exploratorio se realizó en el programa Factor (Lorenzo-Seva & Ferrando, 2013). La adecuación del análisis factorial se calculó a través de la medida KMO de Kaiser y de esfericidad de Bartlett. El método de estimación de factores utilizado fue el de mínimos cuadrados no ponderados. La selección del número de factores se realizó mediante el procedimiento de análisis paralelo (AP; Horn, 1965). La rotación factorial fue oblicua y se utilizó el criterio promin (Lorenzo-Seva, 1999). La confiabilidad se determinó mediante el coeficiente alfa de Cronbach. La validez de criterio se estableció mediante la comparación de la media del puntaje total y la del puntaje por ítem entre el grupo de niños que presentaron dificultades en la alimentación y en el grupo que no presentó dificultades según el cuestionario.

Resultados

Descripción de los ítems individuales y el puntaje total

Doscientos treinta niños de la muestra no presentaron dificultades en la alimentación, mientras que dentro del grupo de los que sí presentaron dificultades se discriminó a dieciocho niños con dificultades leves, ocho con dificultades moderadas y siete con dificultades severas.

En la muestra total, el puntaje medio para los 14 ítems va desde 1.45 a 3.10. Estos valores son consistentes con los valores promedio obtenidos en la muestra normativa de Canadá, así como en los estudios de validación en los Países Bajos y Tailandia. El análisis descriptivo también puso en evidencia que los puntajes por ítem, así como el puntaje total, se distribuían claramente de forma positiva (ver Tabla 1).

Se hallaron diferencias significativas al comparar la media de los puntajes por ítem y del puntaje total en el grupo de niños sin dificultades

Tabla 1

Media de los ítems individuales y del puntaje total de la EADAN para la muestra total (N = 263).

Ítems EADAN	Media	Desv. típ.	Asimetría
Ítem 1	2.67	1.62	0.79
Ítem 2	2.39	1.91	1.04
Ítem 3	2.23	1.57	1.14
Ítem 4	3.10	2.53	0.64
Ítem 5	2.66	1.29	1.14
Ítem 6	2.59	1.70	0.72
Ítem 7	1.87	1.60	2.01
Ítem 8	1.65	1.46	2.44
Ítem 9	2.62	2.10	0.96
Ítem 10	1.85	1.69	2.05
Ítem 11	1.45	1.10	2.68
Ítem 12	1.58	1.26	2.61
Ítem 13	1.60	1.37	2.42
Ítem 14	1.71	1.47	2.13
Puntaje total	30.04	12.30	1.16

des y en el de aquellos con dificultades. Se observa que el grupo de niños con dificultades obtuvo puntajes significativamente superiores en todos los ítems de la escala. El ítem 5 es el único en el que la diferencia entre los dos grupos no resulta tan significativa. La media del puntaje total en el grupo de niños con dificultades duplica el valor del grupo de niños sin dificultades (ver Tabla 2).

Respecto a los aspectos sociodemográficos, en los resultados se puede observar que las madres del grupo de niños que presentaron dificultades poseen en su mayoría estudios secundarios completos (48%), en menor medida estudios terciarios o universitarios completos (21%) y estudios secundarios incompletos (18%). La gran mayoría de estas madres eran amas de casa (46%) mientras que el resto tenía trabajos inestables (21%) o un trabajo estable (33%). No se obtuvie-

Tabla 2
Media de los ítems individuales y del puntaje total de la EADAN para los grupos sin dificultades y con dificultades.

Ítems EADAN	Sin Dificultades (n = 230)		Con Dificultades (n = 33)	
	Media	DE	Media	DE
Ítem 1	2.34	1.36	4.94	1.49
Ítem 2	2.12	1.73	4.27	2.08
Ítem 3	1.94	1.31	4.21	1.79
Ítem 4	2.71	2.37	5.82	1.89
Ítem 5	2.62	1.21	2.94	1.73
Ítem 6	2.37	1.54	4.15	1.97
Ítem 7	1.68	1.41	3.21	2.13
Ítem 8	1.50	1.21	2.73	2.40
Ítem 9	2.24	1.84	5.30	1.79
Ítem 10	1.45	1.13	4.64	2.21
Ítem 11	1.35	0.96	2.15	1.66
Ítem 12	1.45	1.08	2.52	1.88
Ítem 13	1.30	0.92	3.64	2.08
Ítem 14	1.50	1.24	3.15	2.01
Puntaje total	26.53	8.06	54.48	8.28

ron diferencias significativas en relación con el sexo del niño, ya que un 58% eran varones y un 42% eran mujeres.

En cuanto a los antecedentes neonatales de los niños, se encontró que la gran mayoría nació a término (23 niños, 70%), en segundo lugar se encontraron los prematuros moderados (seis niños, 18%) y en igual cantidad los muy prematuros y los extremos (dos niños respectivamente, 6%). El peso al nacer tampoco parecería influir demasiado en este grupo, ya que el 70% tuvo un peso adecuado, mientras que el 12% tuvo muy bajo peso, y el 9% de los niños con dificultades nacieron con bajo peso y extremo bajo peso respectivamente. La mayoría de estos infantes no permanecieron internados en la UCIN al nacer (20 niños, 60%),

no fueron alimentados mediante sonda orogástrica (24 niños, 73%), ni requirieron asistencia respiratoria (24 niños, 73%). Se observa un leve predominio de niños que asisten a la guardería o a la escuela (17 niños, 52%). La mayoría de los niños come junto a la familia (26 niños, 79%) y no poseen antecedentes familiares de trastornos en la alimentación (24 niños, 73%).

Al dividir la muestra en grupos de edad se puede notar que el puntaje medio obtenido en la escala aumenta considerablemente desde los 6 meses hasta los 4 años. De igual manera, también se incrementa la desviación estándar, lo que indica que a medida que los niños crecen, las diferencias inter-individuales son mayores. Si bien la media del puntaje total alcanza su pico a los 4 años, luego desciende levemente entre los 4 años 1 mes y los 5 años para volver incrementarse hasta los 6 años y 11 meses.

Por otra parte, se puede advertir que en el grupo de niños que presentaron dificultades en la alimentación, la mayor concentración se obtuvo en el rango de 6 meses a 1 año (ocho niños), entre los 2 años y 1 mes - 3 años (seis niños) y entre los 4 años y 1 mes - 5 años (seis niños). Esto podría indicar que en la muestra estudiada la mayoría de los infantes presentaron dificultades en el periodo de introducción de los alimentos semisólidos. Sin embargo, al considerar la media del puntaje total obtenido en la EADAN, se observa que los valores más elevados se encuentran en el rango 3 años y 1 mes - 4 años, seguido por el valor medio del rango 4 años y 1 mes - 5 años (ver Tabla 3).

Diferencias entre los niños nacidos prematuros y los nacidos a término

Se realizó una comparación de los puntajes obtenidos por los niños nacidos prematuros y los niños nacidos a término. Se encontró que el 12%

Tabla 3

Media del puntaje total según grupos de edad en la muestra total y en los niños con dificultades en la alimentación.

Grupos (años y meses)	Media	Desv. típ.	N = 263
6 meses a 1 año	26.77	11.80	74
1 año y 1 mes a 2 años	27.36	8.67	45
2 años y 1 mes a 3 años	32.45	12.21	47
3 años y 1 mes a 4 años	35.18	14.74	28
4 años y 1 mes a 5 años	30.74	15.42	27
5 años y 1 mes a 6 años	31.50	13.50	16
6 años y 1 mes a 6 años y 11 meses	33.39	9.28	23

Niños con dificultades en la alimentación	Media	Desv. típ.	n = 33
6 meses a 1 año	53.38	6.61	8
1 año y 1 mes a 2 años	48	1.41	2
2 años y 1 mes a 3 años	55	9.63	6
3 años y 1 mes a 4 años	59.80	13.88	5
4 años y 1 mes a 5 años	55.67	6.91	6
5 años y 1 mes a 6 años	48.75	1.50	4
6 años y 1 mes a 6 años y 11 meses	50.50	2.12	2

de los niños que nacieron prematuros tuvieron dificultades en la alimentación según el instrumento, mientras que estas dificultades aparecieron en el 13% de los niños nacidos a término.

En el grupo de prematuros extremos el rango de EG al nacer abarcó entre 25 y 27 semanas de gestación, con un valor medio de 26.25 semanas. La EG media en los niños nacidos muy prematuros fue de 29.76 con un rango de edad de entre 28 y 31 semanas. En los prematuros moderados la media fue de 34.42 con un rango de edad entre 32 y 36 semanas. Finalmente, en el grupo de nacidos a término la EG media fue de 39.10 semanas con un rango de edad entre 37 y 42 semanas de gestación.

Se observa que el grupo de los prematuros extremos obtuvo el puntaje medio total más alto de la muestra con un valor de 32 puntos. También se puede notar que los prematuros extremos tuvieron el puntaje medio por ítem más alto en siete de las variables estudiadas, particularmente en aquellas que se refieren a conductas como negarse a comer, irregularidad al comer, mal comportamiento durante las comidas, guardar la comida en la boca sin tragarla; así como también en los ítems que indicarían dificultades en las habilidades oromotoras y pobre crecimiento. En cambio, se observa que en este grupo la alimentación no afectaría las relaciones con los padres y la familia en su totalidad, probablemente debido a la atención especial que recibieron durante la internación y en el seguimiento.

Los niños considerados como muy prematuros obtuvieron el puntaje medio por ítem más alto en cuatro de las variables indagadas, en especial en aquellas que remiten a tener arcadas, escupir o vomitar con cierto tipo de comidas y la necesidad de entretenerlo o perseguirlo para que coma, así como la necesidad de forzarlo a hacerlo. Finalmente, el grupo de prematuros moderados y los nacidos a término obtuvieron los puntajes medios más altos sólo en dos de los ítems del cuestionario (ver Tabla 4).

Tabla 4

Comparación entre nacidos prematuros y nacidos a término respecto de los puntajes medios por ítem y total de la EADAN.

Ítems	Prematuros extremos n = 8		Muy Prematuros n = 21		Prematuros Moderados n = 57		Nacidos a término n = 177	
	Media	Desv. típ.	Media	Desv. típ.	Media	Desv. típ.	Media	Desv. típ.
Ítem 1	2.63	2.387	2.62	1.465	2.30	1.476	2.80	1.649
Ítem 2	2.75	1.909	2.14	1.957	2.33	2.099	2.42	1.867
Ítem 3	1.88	1.246	1.86	1.153	2.39	1.770	2.24	1.559
Ítem 4	4.25	3.012	2.76	2.587	2.53	2.315	3.27	2.562
Ítem 5	3.25	1.165	2.86	1.621	2.65	1.395	2.62	1.220
Ítem 6	2.88	2.167	2.67	1.653	2.16	1.590	2.71	1.720
Ítem 7	2.13	2.232	2.57	1.777	1.86	1.586	1.78	1.546
Ítem 8	2.13	2.100	1.76	1.609	1.70	1.488	1.60	1.423
Ítem 9	1.88	1.642	2.67	2.033	2.53	2.292	2.68	2.070
Ítem 10	1.38	1.061	2.14	2.242	1.82	1.794	1.84	1.609
Ítem 11	1.75	1.753	1.14	.359	1.51	1.325	1.46	1.050
Ítem 12	2.63	1.923	1.81	1.209	1.68	1.572	1.47	1.093
Ítem 13	1.50	1.414	1.24	.768	1.67	1.574	1.62	1.360
Ítem 14	1.00	.000	1.86	1.740	1.49	1.151	1.79	1.551
Puntaje total	32.00	12.189	30.10	11.717	28.49	13.505	30.45	12.026

Propiedades psicométricas de la escala

Se analizó la confiabilidad de la escala mediante el estadístico alfa de Cronbach, a partir de lo cual se obtuvo que la misma es buena con un valor $\alpha = .79$. Se realizó un análisis factorial exploratorio en el programa Factor (Lorenzo-Seva & Ferrando, 2013). El estadístico KMO arrojó un valor de .85, por lo que se considera satisfactorio, y en la prueba de Bartlett la significación fue de $p < .001$. A partir de estos estadísticos se pudo comprobar la adecuación de los datos al análisis factorial, por lo que resulta apropiado realizarlo. El método de extracción de factores utilizado fue el de mínimos cuadrados no ponderados. Se seleccionó el número de factores mediante el proce-

dimiento de análisis paralelo (AP), que arrojó dos factores que representan el 49.4% de la varianza. Se retuvieron aquellos ítems cuya saturación fue superior a .30 (ver Tabla 5). La correlación entre los dos factores es alta, de .76.

La estructura factorial subyacente permite establecer que el primer factor agrupa los ítems referidos a *causas y consecuencias de las dificultades en la alimentación*; entre ellas se puede mencionar el poco apetito, la negativa a comer, las habilidades oromotoras y orosensoriales deficientes y el pobre crecimiento del niño. El segundo factor abarca los ítems que se refieren a *preocupaciones parentales y estrategias utilizadas frente a las dificultades en la alimentación*, tales como mal comportamiento durante las comidas,

Tabla 5

Factores obtenidos a partir del método mínimos cuadrados no ponderados y promin. Los valores mayores a .30 se encuentran resaltados.

	Ítems EADAN	Factor 1	Factor 2
1:	¿Cómo describe las comidas con su hijo?	.156	.671
6:	¿Cómo se comporta su hijo en las comidas?	-.386	.955
10:	¿Tiene que forzarlo a comer/beber?	.254	.589
13:	¿Cómo afecta la alimentación su relación con su hijo?	.037	.702
9:	¿Tiene que entretenerlo o perseguirlo para que coma?	.242	.555
14:	¿Qué efecto tiene la alimentación de su hijo en las relaciones familiares?	.064	.610
4:	¿En qué momento de la comida su hijo se niega a comer?	.372	.343
5:	¿Cuánto tiempo le lleva comer a su hijo?	-.257	.419
2:	¿Está preocupada por la alimentación de su hijo?	.759	-.079
3:	¿Cuánto apetito tiene su hijo?	.622	.105
11:	¿Cómo mastica/succiona su hijo?	.726	-.287
12:	¿Cómo describe el crecimiento de su hijo?	.536	-.027
7:	¿Tiene arcadas, escupe o vomita con ciertos tipos de comidas?	.470	-.002
8:	¿Guarda la comida en la boca sin tragársela?	.308	.136

influencia en las relaciones familiares, estrategias como forzar al niño a comer, entretenerlo o perseguirlo. Se puede observar que en el ítem 4 los valores para cada factor son cercanos, por lo que se infiere que este ítem es estadísticamente débil y no tendría un peso muy significativo en la constitución de los factores.

Discusión

En el presente artículo se exponen las características psicométricas de la EADAN, que consiste en la adaptación argentina y la validación en nuestra población de la MCH-FS. Los resultados expuestos sustentan la utilización de la EADAN como un instrumento para la detección de dificultades en la alimentación en niños. El coeficiente alfa de Cronbach fue satisfactorio para

la EADAN, lo que indica una buena consistencia interna.

En nuestra investigación se obtuvieron dos factores que subyacen a las variables estudiadas. El Factor 1 se refiere a las *causas y consecuencias de las dificultades en la alimentación* y el segundo factor a las *preocupaciones parentales y estrategias utilizadas frente a las dificultades en la alimentación*. Se puede advertir que el ítem 4 sería estadísticamente débil ya que no presenta diferencias tan marcadas entre los valores para un factor y para otro, por lo que sería necesario revisar en estudios posteriores si es necesario retirarlo o modificarlo para ser incorporado en el instrumento.

Se pueden observar diferencias significativas tanto en los puntajes medios por ítem como en el puntaje medio total obtenido en el grupo de niños con dificultades en la alimentación y el grupo de niños sin dificultades. Esto evidencia que

en la muestra argentina los valores del grupo con dificultades son claramente superiores a los valores del grupo sin dificultades, lo que coincide con los resultados de los estudios realizados en Canadá (Ramsay et al., 2011) y en Tailandia (Benjasuwantep et al., 2015).

Al comparar los puntajes medios de los ítems obtenidos por los grupos con y sin dificultades, se puede observar que ciertos comportamientos considerados problemáticos en torno a la alimentación se encuentran presentes en ambos grupos. Esto coincide con lo señalado por Ramsay et al. (2011) respecto de que la diferencia estaría en la frecuencia con la que se presentan estos comportamientos o en la percepción de los padres respecto de la severidad de los mismos y las reacciones ante ellos. En nuestra investigación, la frecuencia de las dificultades en la alimentación es del 12%, inferior a lo que se estima en la literatura (Arvedson, como se citó en Barton et al., 2017) y a los resultados obtenidos en el estudio canadiense (Ramsay et al., 2011).

Al comparar los valores medios por ítem y el valor medio del puntaje total de la investigación argentina con los estudios anteriores se puede notar que, respecto de la muestra normativa de Canadá, los valores argentinos por ítem y el puntaje total son claramente inferiores. Esto podría indicar que las conductas problemáticas son menos frecuentes o menos intensas en nuestra muestra; sin embargo, esto requiere ser estudiado en mayor profundidad. No ocurre lo mismo con el estudio holandés y el tailandés, en los que el puntaje medio es inferior en algunos ítems y superior en otros. Esto podría estar relacionado con las diferencias en el rango de edad de los niños estudiados, ya que en Canadá, al igual que en el presente estudio, se evaluaron niños de entre los 6 meses y los 6 años 11 meses; en cambio, en las otras dos investigaciones el rango abarcó hasta los 4 años de edad.

En la presente investigación se encontró que los niños que fueron prematuros extremos obtuvieron los puntajes medios por ítem y el puntaje total más elevado dentro de la muestra en estudio. Esto corrobora lo señalado en la literatura (Hvelplund et al., 2016) respecto de la EG inferior a 28 semanas como un factor que puede incrementar el riesgo de aparición de dificultades en la alimentación. Por otro lado, es importante mencionar que en esta investigación no se hallaron diferencias significativas en cuanto a la frecuencia con la que se presentaron las dificultades en la alimentación entre niños prematuros y a término (12% vs 13%). Esto coincide con lo señalado por Van Dijk et al. (2016), quienes obtuvieron que la mayoría de los prematuros estudiados en su investigación no presentaban problemas de alimentación o bien presentaban problemas leves. Si bien estos resultados parecen contradecir las numerosas investigaciones que señalan a la prematuridad como un factor de riesgo para este tipo de problemas, sería necesario considerar la posibilidad de ampliar la cantidad de prematuros evaluados para profundizar este análisis. Sin embargo, resulta evidente que a medida que disminuye la edad gestacional al nacer aumentan los puntajes medios en los ítems del cuestionario. Los resultados permiten señalar que, si bien el puntaje total del instrumento es relevante, resulta fundamental considerar los valores obtenidos en los ítems individuales, ya que esto permitirá delinear el tratamiento más oportuno en cada caso.

Es importante señalar que los resultados obtenidos en la presente investigación coinciden con los hallazgos de Van Dijk et al. (2011) respecto del incremento de la media del puntaje total a medida que aumenta la edad de los niños. Esto se puede notar en la muestra general, en la que la media va incrementándose a partir de los 6 meses para alcanzar su pico a los 4 años. De igual manera, en el grupo de 33 niños que tuvieron di-

ficultades en la alimentación, el valor medio más elevado se obtuvo entre los 3 años y 1 mes y los 4 años, lo que coincide con lo obtenido por los autores holandeses. Esto podría estar relacionado con las características de la etapa por la que atraviesan los niños, caracterizada por un mayor oposiciónismo y berrinches.

Una limitación del presente estudio es el tamaño de la muestra, que no permitiría realizar generalizaciones a la población respecto de la frecuencia de los problemas de alimentación en los niños. Otra limitación es que los cuestionarios fueron completados por los padres mientras esperaban con sus hijos para la consulta pediátrica o una vez finalizada la misma. Esto supone que los padres tenían a sus hijos en brazos, en ocasiones los alimentaban o estaban pendientes de que no se alejaran, lo que podría afectar la atención y la información brindada en el cuestionario. Tal vez sería conveniente ofrecer a los padres que completen el cuestionario en el hogar y luego lo devuelvan o enviárselos por correo electrónico.

A partir de la presente investigación se puede concluir que la EADAN es un instrumento de cribado confiable, válido para la detección de dificultades en la alimentación en niños de entre 6 meses y 6 años y 11 meses de la población argentina. Se trata de un instrumento de gran utilidad para su implementación en la consulta pediátrica, tanto en los controles médicos de niños sanos como los de nacidos prematuros. Este instrumento resulta de gran importancia en nuestro país ya que permitiría cubrir la vacancia respecto de instrumentos adaptados y validados que permitan una rápida detección de las dificultades en la alimentación en los niños pequeños. Se trata de un instrumento de fácil administración y fácil comprensión por los padres en nuestra población.

En investigaciones futuras sería recomendable ampliar la muestra tanto en el número de sujetos como en las localidades en las que se ad-

ministre el instrumento a fin de poder desarrollar baremos locales para los profesionales de la salud. La relevancia de continuar investigando en esta dirección surge del conocimiento respecto del impacto negativo que tienen las dificultades en la alimentación en el crecimiento y el desarrollo de los niños, así como en las relaciones familiares. La detección temprana de estas problemáticas contribuiría a mejorar la salud física de los niños, así como los vínculos familiares, ofreciendo una oportunidad para que todos los niños puedan acceder a un crecimiento y desarrollo saludable.

Referencias

- Adamson, M., Morawska, A., & Sanders, M. R. (2013). Childhood feeding difficulties: A randomized controlled trial of a group-based parenting intervention. *Journal of Developmental & Behavioral Pediatrics, 34*(5), 293-302. doi: [10.1097/DBP.0b013e3182961a38](https://doi.org/10.1097/DBP.0b013e3182961a38)
- Barton, C., Bickell, M., & Fucile, S. (2017). Pediatric oral motor feeding assessments: A systematic review. *Physical & Occupational Therapy in Pediatrics, 38*(2), 190-209. doi: [10.1080/01942638.2017.1290734](https://doi.org/10.1080/01942638.2017.1290734)
- Benjasuwantep, B., Rattanamongkolgul, S., & Ramsay, M. (2015). The Thai version of the Montreal Children's Hospital Feeding Scale (MCH-FS): Psychometric properties. *Journal of Medical Association of Thailand, 98*(2), 163-169. Recuperado de <http://www.jm-atonline.com/index.php/jmat>
- Blisseth, J. (2011). Relationships between parenting style, feeding style and feeding practices and fruit and vegetable consumption in early childhood. *Appetite, 57*(3), 826-831. doi: [10.1016/j.appet.2011.05.318](https://doi.org/10.1016/j.appet.2011.05.318)
- Fernández-Gallardo, M. A., Rojas-Contreras, D. P., & Vargas-Keith, J. F. (2017). Development of feeding skills in preterm infants: A critical literature review. *Revista CEFAC, 19*(4), 539-550. doi: [10.1590/1982-021620171946417](https://doi.org/10.1590/1982-021620171946417)

- Goh, D. & Jacob, A. (2012). Perception of picky eating among children in Singapore and its impact on caregivers: A questionnaire survey. *Asia Pacific Family Medicine*, 11(5), 1-8. doi: [10.1186/1447-056X-11-5](https://doi.org/10.1186/1447-056X-11-5)
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30(2), 179-185. doi: [10.1007/bf02289447](https://doi.org/10.1007/bf02289447)
- Hvelplund, C., Hansen, B. M., Vinkel-Koch, S., Andersson, M., & Skovgaard, A. M. (2016). Perinatal risk factors for feeding and eating disorders in children aged 0 to 3 years. *Pediatrics*, 137(2). doi: [10.1542/peds.2015-2575](https://doi.org/10.1542/peds.2015-2575)
- IBM Corp. (2012). IBM SPSS Statistics for Windows, Version 21.0. [software de cómputo]. Armonk, NY: IBM Corp.
- Johnson, R., & Harris, G. (2004). A preliminary study of the predictors of feeding problems in late infancy. *Journal of Reproductive and Infant Psychology*, 22(3), 183-188. doi: [10.1080/02646830410001723760](https://doi.org/10.1080/02646830410001723760)
- Lindberg, L., Bohlin, G., & Hagekull, B. (1991). Early feeding problems in a normal population. *International Journal of Eating Disorders*, 10(4), 395-405. doi: [10.1002/1098-108x\(199107\)10:4<395::aid-eat2260100404>3.0.co;2-a](https://doi.org/10.1002/1098-108x(199107)10:4<395::aid-eat2260100404>3.0.co;2-a)
- Lorenzo-Seva, U. (1999). Promin: A method for oblique factor rotation. *Multivariate Behavioral Research*, 34(3), 347-356. doi: [10.1207/S15327906MBR3403_3](https://doi.org/10.1207/S15327906MBR3403_3)
- Lorenzo-Seva, U., & Ferrando, P. J. (2013). FACTOR 9.2: A comprehensive program for fitting exploratory and semiconfirmatory factor analysis and IRT models. *Applied Psychological Measurement*, 37(6), 497-498. doi: [10.1177/0146621613487794](https://doi.org/10.1177/0146621613487794)
- McDermott, B. M., Mamun, A. A., Najman, J. M., Williams, G. M., O'Callaghan, M. J., & Bor, W. (2008). Preschool children perceived by mothers as irregular eaters: Physical and psychosocial predictors from a birth cohort study. *Journal of Developmental and Behavioral Pediatrics*, 29(3), 197-205. doi: [10.1097/DBP.0b013e318163c388](https://doi.org/10.1097/DBP.0b013e318163c388)
- Ramsay, M. (2013). Feeding skill, appetite and feeding behaviours of infants and young children and their impact on growth and psychosocial development. En R. E. Tremblay, R. G. Barr & R. DeV. Peters (Eds.), *Encyclopedia on Early Childhood Development* [online] (1-8). Montreal, Quebec: Centre of Excellence for Early Childhood Development. Recuperado de <http://www.child-encyclopedia.com>
- Ramsay, M., Martel, C., Porporino, M., & Zygmuntowicz, C. (2011). The Montreal Children's Hospital Feeding Scale: A brief bilingual screening tool for identifying feeding problems. *Paediatrics and Child Health*, 16(3), 147-e17. doi: [10.1093/pch/16.3.147](https://doi.org/10.1093/pch/16.3.147)
- Rommel, N., De Meyer, A. M., Feenstra, L., & Veerman-Wauters, G. (2003). The complexity of feeding problems in 700 infants and young children presenting to a tertiary care institution. *Journal of Pediatric Gastroenterology and Nutrition*, 37(1), 75-84. doi: [10.1097/00005176-200307000-00014](https://doi.org/10.1097/00005176-200307000-00014)
- Sanders, M. R., Patel, R. K., Le Grice, B., & Shepherd, R. W. (1993). Children with persistent feeding difficulties: An observational analysis of the feeding interactions of problem and non-problem eaters. *Health Psychology*, 12(1), 64-73. doi: [10.1037//0278-6133.12.1.64](https://doi.org/10.1037//0278-6133.12.1.64)
- Saxton, J., Carnell, S., van Jaarsveld, C. H. M., & Wardle, J. (2009). Maternal education is associated with feeding style. *Journal of the American Dietetic Association*, 109(5), 894-898. doi: [10.1016/j.jada.2009.02.010](https://doi.org/10.1016/j.jada.2009.02.010)
- Van Dijk, M., Bruinsma, E., & Hauser, M. P. (2016). The relation between child feeding problems as measured by parental report and mealtime behavior observation: A pilot study. *Appetite*, 99, 262-267. doi: [10.1016/j.appet.2016.01.026](https://doi.org/10.1016/j.appet.2016.01.026)
- Van Dijk, M., Timmerman, M. E., Martel, C., & Ramsay, M. (2011). Towards the development of a Dutch screening instrument for the detection of feeding problems in young children. *Netherlands Journal of Psychology*, 66(4), 112-119. Recuperado de [https://www.rug.nl/research/portal/journals/netherlands-journal-of-psychology\(11473cbc-a542-47fc-acbb-736cd32ce42c\)/publications.html](https://www.rug.nl/research/portal/journals/netherlands-journal-of-psychology(11473cbc-a542-47fc-acbb-736cd32ce42c)/publications.html)

Proceso de construcción de pruebas educativas: El caso de la Prueba de Habilidades Cuantitativas

Educational tests development process: The case of the Quantitative Abilities Test

Luis Rojas-Torres * ¹, Graciela Ordóñez-Gutiérrez ²

1 - Universidad de Costa Rica, Costa Rica.

Introducción
Construcción
de una prueba
Etapas previas
PHC 2018
Discusión
Referencias

Recibido: 08/05/2019 Revisado: 17/05/2019 Aceptado: 24/05/2019

Resumen

La finalidad de este artículo es brindar una guía teórica y práctica de cómo construir una prueba educativa. En la primera parte del artículo se presenta una exposición detallada de las etapas que se deben seguir para construir una prueba educativa escrita. En la segunda parte, se muestra cómo se aplicó cada una de estas etapas a la construcción de la Prueba de Habilidades Cuantitativas de la Universidad de Costa Rica, una prueba educativa con ítems de selección única que utiliza el modelo de medición de Teoría de Respuesta al Ítem de dos parámetros. A partir de la exposición, se concluye que la construcción de pruebas es un proceso riguroso, por lo que una elaboración deficiente provocaría una generación de inferencias erróneas de las habilidades de los sujetos.

Palabras clave: *prueba educativa, construcción de pruebas, evidencias de validez, Prueba de Habilidades Cuantitativas, Teoría de Respuesta al Ítem*

Summary

The goal of this paper is to present a theoretical and practical guide on how to develop an educational test. In the first part of this paper, a detailed explanation of each stage used in development of a written educational test is presented. In the second part, it is shown how this process was applied to the development of the Quantitative Abilities Test of the University of Costa Rica, which is an educational test with multiple choice items that uses the two parameter measurement model of the Item Response Theory. From this exposition, it is concluded that the test development is a rigorous process in which even a little mistake can cause wrong inferences about subjects' abilities.

Keywords: *educational test, test development, validity evidences, Quantitative Abilities Test, Item Response Theory*

*Correspondencia a: Luis Rojas-Torres, E-mail: luismiguel.rojas@ucr.ac.cr

Cómo citar este artículo: Rojas-Torres, L., & Ordóñez-Gutiérrez, G. (2019). Proceso de construcción de pruebas educativas: El caso de la Prueba de Habilidades Cuantitativas. *Revista Evaluar*, 19(2), 15-29. Recuperado de <https://revistas.unc.edu.ar/index.php/revaluar/index>

Introducción

Las pruebas o tests son instrumentos o dispositivos de evaluación de un dominio específico que permiten medir el grado de acierto de las respuestas que los sujetos otorgan a un conjunto de preguntas (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education [AERA, APA, & NCME], 2014; Castillo-Arredondo & Cabrerizo-Diago, 2010); además, son los instrumentos de evaluación más utilizados en los sistemas educativos, ya que permiten la valoración de componentes educativos de manera directa; asimismo, los tests posibilitan calificaciones estandarizadas a un corto plazo.

En particular, las pruebas educativas son aquellas que evalúan elementos relacionados con el proceso de enseñanza y aprendizaje de los estudiantes, como el dominio de un contenido o procesos de razonamiento en los que se recurre al contenido aprendido. Los principales propósitos con que se usan estas pruebas son: a) para hacer inferencias de los resultados educativos estudiantiles; b) para tomar decisiones con respecto a los estudiantes (certificaciones, diagnóstico, selección o ingreso a programas educativos); c) para realizar inferencias sobre los procesos de enseñanza y aprendizaje (AERA, APA, & NCME, 2014) y d) para evaluar la calidad de los sistemas educativos en pro de la mejora de estos (Castillo-Arredondo & Cabrerizo-Diago, 2010; Tiana, 1996). A pesar de que las pruebas educativas son ampliamente utilizadas en diversas áreas de la educación, su construcción muchas veces se realiza sin la rigurosidad técnica demandada por las múltiples instancias especializadas en evaluación y medición educativa y las consideraciones establecidas en los múltiples manuales que orientan su desarrollo. Por este motivo, este artículo tiene como objetivo brindar una guía teórica y práctica

de cómo construir una prueba educativa escrita. Para este objetivo se presentan las etapas que rigen la construcción de una prueba educativa, de acuerdo con lo establecido en la literatura; luego, se presenta un ejemplo de la aplicación de esta guía en la construcción de la Prueba de Habilidades Cuantitativas (PHC) de la Universidad de Costa Rica (UCR), utilizada en el proceso de admisión del año 2018.

Es importante mencionar que el proceso que se describe en este artículo es aplicable tanto para pruebas estandarizadas como para “pruebas de aula” (aquellas que utilizan los docentes para la evaluación del avance de los estudiantes durante los ciclos lectivos). No obstante, la descripción de los procesos de este documento se concentra en las evaluaciones estandarizadas, las cuales disponen de un tiempo de construcción más amplio que el que tienen las pruebas de aula. La recomendación para el docente el aula es que procure acercarse lo más que pueda a los estándares descritos en este documento.

Construcción de una prueba educativa

Son varios los investigadores y especialistas en construcción de pruebas (Downing, 2006; Embretson, 2017; Ferrara, Lai, Reilly, & Nichols, 2017; Kane, 2013; Muñoz & Fonseca-Pedrero, 2019; Villareal-Galera, Alfaro-Rojas, & Brizuela-Rodríguez, 2015) que indican que para elaborar un test se requiere una secuencia de etapas, pasos o fases, de tal manera que se alcancen evidencias de validez y confiabilidad sobre las puntuaciones obtenidas en estos. A continuación se presenta una explicación de las etapas fundamentales para la construcción de tests.

Plan general

La primera etapa en la construcción de una prueba es definir el *qué* y *para qué* se quiere medir. El *qué* medir determina el constructo, lo cual implica efectuar un abordaje teórico de este y una explicación de cómo la medición del constructo permite el alcance del objetivo establecido o el propósito de la medición. Hay objetivos que se explican directamente, como la obtención de los niveles de conocimiento en el constructo de los miembros de una población, pero hay otros que son más complejos como es el uso de una prueba de razonamiento con figuras para la selección de los estudiantes de una carrera universitaria. El *para qué* medir implica establecer una explicación detallada y precisa de las razones por las cuales se elabora el test (Muñiz & Fonseca-Pedrero, 2019). Además, se debe especificar el contexto en el cual se aplicará la prueba, el cual incluye la población a la que se medirá y las circunstancias de la aplicación. También es importante tomar en cuenta qué decisiones se tomarán con respecto a las personas de acuerdo con las puntuaciones obtenidas, ya que la calificación en un test puede servir para propósitos varios como seleccionar, diagnosticar, clasificar, entre otros.

Las respuestas al *qué* y *para qué* medir determinarán todos los elementos que se considerarán en la construcción. De esta forma los reactivos de la prueba serán de respuesta cerrada si se pretende evaluar el producto final de un proceso, o serán de respuesta abierta si es para evaluar todo el proceso (Castillo-Arredondo & Cabrerizo-Diago, 2010; Mateo & Martínez, 2008). Además, el modelo de interpretación de los puntajes será con base en normas si se pretende comparar a los sujetos con el resto de la población, o será con base en criterios si lo que se quiere es establecer estándares de dominio (Martínez-Arias, Hernández-Lloreda, & Hernández-Lloreda, 2006). Por otro lado, el

modelo de medición será la Teoría Clásica de los Tests (TCT), la Teoría de Respuesta de los Ítems (TRI) u otro, según las propiedades del modelo que beneficien al objetivo de medición (Muñiz & Fonseca-Pedrero, 2019). Por ejemplo, si la finalidad es maximizar la discriminación en un nivel de habilidad se puede recurrir a la TRI; pero, si se quiere medir un constructo en una población pequeña se puede recurrir a la TCT.

Cabe resaltar que el plan general es indispensable para alcanzar un grado aceptable de validez de los usos de las puntuaciones de un test, donde *validez* se entiende como “el grado en que la evidencia empírica y las justificaciones teóricas apoyan la pertinencia de las acciones e interpretaciones de las puntuaciones de las pruebas” (Messick, 1989, p. 6). Si una prueba se construye sin tener en cuenta cuáles interpretaciones se harán con sus puntajes o qué usos se les darán a sus puntuaciones, difícilmente tendrá interpretaciones o usos válidos desde la perspectiva de la rigurosidad de la evaluación y medición educativa.

Definición del contenido

En esta etapa se establece cuáles son los elementos, componentes o dimensiones del constructo a evaluar. La definición de estos componentes se realiza a partir de la revisión teórica sobre el constructo o mediante la construcción de una teoría acerca del mismo (Embretson, 2017; Muñiz & Fonseca-Pedrero, 2019). Esto con la finalidad de constituir una definición operativa del constructo y lograr obtener medidas de manera empírica. En esta etapa se empieza el desarrollo de una de las primeras fuentes de evidencias de validez de los usos de las puntuaciones: evidencias de validez basadas en el contenido, las cuales buscan que todos los elementos relevantes del constructo sean considerados (AERA, APA, & NCME, 2014).

Especificaciones del test

En esta etapa se define cómo deben ser las características del instrumento con el que se evaluará el constructo pretendido, por lo que se debe construir una tabla de especificaciones, ya que esta es un elemento indispensable a la hora de elaborar los ítems del test. En la tabla se asigna el puntaje que se debe otorgar a cada combinación de categorías de distintos aspectos del test como procesos, contenidos, dificultades, entre otros. El uso riguroso de la tabla de especificaciones permitirá que la prueba presente evidencias de validez basadas en el contenido.

Además, se debe decidir cuál será el formato de los ítems con los que se evaluarán los elementos de la tabla de especificaciones, esto es: el tipo de ítem, la longitud y el tipo de alternativas que se va a utilizar (Martínez-Arias et al., 2006; Muñiz & Fonseca-Pedrero, 2019). Igualmente, se debe dilucidar qué tiempo será necesario para resolver la prueba, los materiales que se utilizarán para la evaluación y los horarios en que se aplicará la prueba. Cada uno de estos elementos se selecciona considerando la finalidad de emplear una prueba. Por otro lado, esta etapa es la base para alcanzar evidencias de validez basadas en la estructura interna, que son aquellas evidencias de que las dimensiones establecidas en la definición del constructo se reproducen en los datos de la prueba.

Construcción de los ítems

De acuerdo con Muñiz y Fonseca-Pedrero (2019), la construcción de los ítems constituye una de las fases más importantes en la confección de un instrumento de medición, particularmente en la elaboración de una prueba educativa; puesto que los reactivos son los que conforman

el instrumento. Así, una construcción deficiente incidirá en las propiedades métricas del test, lo que repercutirá mucho en las inferencias que se realicen sobre las puntuaciones que se obtengan. En este sentido, para la elaboración de los ítems es indispensable capacitar a las personas que realizarán esta labor. Esto demanda generar el perfil de los constructores, contactar a personas con ese perfil dispuestas a colaborar en la construcción, brindar una capacitación detallada de los elementos que se desean evaluar en los ítems, asignar tareas específicas de cómo y qué incluir en la elaboración de los reactivos. Luego se debe evaluar la construcción. Con base en esta evaluación se selecciona a los mejores constructores y se procede a la construcción de los ítems. Para esta etapa es recomendable otorgar y asignar los elementos particulares de la tabla de especificaciones a cada constructor.

Una vez construidos los reactivos, estos deben ser evaluados por un grupo de expertos en el constructo a medir. En esta valoración se debe analizar si los elementos pretendidos del constructo están considerados y, también, se debe indagar si no hay fuentes de varianza irrelevante al constructo, es decir, que en los ítems no se evalúen elementos que no forman parte del constructo en cuestión (Messick, 1989). Además, se debe analizar si hay fuentes de dificultad diferenciales por grupos relevantes de población. Por ejemplo, en una prueba realizada por dos culturas distintas no se deben agregar contextos familiares para una sola de ellas. En este mismo sentido, se solicita a los expertos que juzguen si los ítems cumplen con los principios básicos que deben regir la construcción de ítems, estos son de acuerdo con Muñiz y Fonseca-Pedrero (2019): representatividad, relevancia, diversidad, claridad y sencillez. Los reactivos que se aprueban con este juzgamiento pueden ser utilizados en el ensamblaje de la prueba. Es importante resaltar que en esta etapa

se rechazan muchos reactivos, sobre todo cuando los constructores son novatos, debido a esto es importante efectuar una construcción de al menos el doble de la cantidad pretendida.

Estudio piloto de los ítems

Luego de la construcción de los reactivos, se procede a ensamblar uno o varios formularios de aplicación según las características establecidas en la tabla de especificaciones. Estos formularios serán aplicados a una población con características semejantes a la población a la que va dirigida la prueba con el fin de obtener una aproximación de las propiedades psicométricas del instrumento en la población meta, lo cual permitirá determinar cuáles ítems son aptos para ser utilizados en el instrumento final (Castillo-Arredondo & Cabreri-zo-Diago, 2010; Mateo & Martínez, 2008; Muñiz & Fonseca-Pedrero, 2019; Villarreal-Galera et al., 2015). En esta etapa se debe procurar mantener las mismas condiciones de administración de la prueba. En cuanto a los análisis estadísticos del test, se deben realizar los requeridos según el modelo de medición seleccionado inicialmente. La explicación de la aplicación y el análisis de ítems, se especifica en las etapas de la construcción de la prueba definitiva.

Con los datos de esta aplicación piloto, se pueden indagar evidencias de validez basadas en la relación con otras variables (Embretson, 2017; Martínez-Arias et al., 2006; Villarreal-Galera et al., 2015), que es la comprobación de relaciones teóricas del constructo pretendido con variables externas. Esto se refiere a los patrones de relación de las puntuaciones de la prueba con otras puntuaciones de rasgos y criterios empíricos que estén relacionadas con los rasgos representados por la calificación obtenida por los examinados en la prueba; por ejemplo, lugar de procedencia

escolar, medidas de motivación, entre otras. Es importante que la evidencia empírica sobre las relaciones con otras variables sea consistente con los objetivos de medición para respaldar las evidencias de contenido (Embretson, 2017).

Para obtener evidencias de relación con otras variables se requiere construir una base de datos que incluya información relevante sobre los examinados, por ejemplo: cantidad de materias matriculadas, promedios obtenidos en cursos relevantes, etc. Por otro lado, se puede indagar sobre las evidencias de validez basadas en la estructura interna mediante la comprobación de la estructura factorial propuesta (Martínez-Arias et al., 2006; Mateo & Martínez, 2008). Las indagaciones de evidencias de validez desde el pilotaje permitirán que la prueba final presente las evidencias de validez requeridas.

Ensamblaje de la prueba

En esta etapa se seleccionan los ítems que serán utilizados en el examen. Para la selección de los ítems se deben seguir las condiciones establecidas en la tabla de especificaciones. Además, si en alguna de las condiciones establecidas hay un excedente de ítems, lo recomendable es seleccionar los ítems que, según los jueces, sean más pertinentes para la evaluación del constructo pretendido.

Posteriormente, se procede a generar el formulario de examen. Este formulario debe iniciar con unas instrucciones generales que señalen: a) el tiempo requerido para resolver la prueba completa; b) la forma en que debe resolverse; y, c) la estructura de la prueba en cuanto a cantidad de ítems. Luego del ensamblaje, se debe realizar una revisión detallada del formulario para garantizar que los ítems no tengan errores de forma ni de fondo.

Aplicación de la prueba

La planificación de la aplicación de la prueba es la etapa en la que se definen las condiciones necesarias para la administración del test en la población meta. Es por esto que esta etapa demanda efectuar una logística rigurosa para asegurar dichas condiciones. Un punto importante es la capacitación de los aplicadores, quienes son los encargados de administrar el test a los examinados.

La capacitación implica brindar la información sobre las labores importantes en la administración de la prueba; por ejemplo: el resguardo del material, la revisión del aula, la organización de entrada de los postulantes a las instalaciones, la explicación correcta de las instrucciones, la supervisión de la aplicación y la devolución del material. En la capacitación se debe asegurar que los aplicadores puedan brindar las condiciones requeridas para que los sujetos demuestren su verdadero nivel de habilidad en la prueba (Muñiz & Fonseca-Pedrero, 2019). Un error puede alterar el significado de los resultados; por ejemplo, la aplicación de una prueba en un lugar sin iluminación adecuada o muy ruidoso puede producir que varios sujetos no logren tener un buen desempeño en la prueba y llevarlos a no reflejar su verdadera habilidad, lo que implicaría una inadecuada interpretación sobre las puntuaciones y, por ende, los usos de las pruebas tendrían un bajo grado de validez.

Calificación de los ítems

En esta etapa se requiere de una guía de calificación. Los ítems cuyo formato de respuesta es cerrado son los más sencillos de calificar, ya que la guía solo debe contener las respuestas correctas. En el caso de las preguntas de respuesta abierta se requiere de una rúbrica o escala en la que: a)

se esbocen las respuestas a esperar; y, b) se especifique la puntuación por cada una de las partes de las respuestas consideradas correctas (Mateo & Martínez, 2008). Además, las guías deben ser lo más exhaustivas posible con respecto a las distintas formas de resolución. Por otro lado, en la calificación de las preguntas de respuesta abierta se deben desarrollar procesos de equiparación por jueces para que las puntuaciones no estén sesgadas por las diferencias entre la severidad de los jueces. La variación en los criterios de calificación de los jueces es una amenaza a la validez, dado que lleva a que la puntuación considere elementos irrelevantes para la medición.

Una vez calificados los ítems se debe realizar el análisis estadístico de estos, con base en el modelo de medición establecido desde la primera etapa. En este sentido, se analizará la calidad psicométrica de las puntuaciones de cada ítem, por ejemplo discriminación y dificultad, y de la prueba en general según el modelo considerado. En caso de que existan ítems que no cumplan los estándares establecidos por la teoría, estos no deben ser considerados en la calificación final (Ferrara et al., 2017; Martínez-Arias et al., 2006; Muñiz & Fonseca-Pedrero, 2019).

Con el análisis de los ítems, se busca generar evidencias de validez basadas en la estructura interna por medio de la comprobación de la configuración factorial establecida en la teoría. También se buscan evidencias de precisión (o confiabilidad), es decir, que las puntuaciones de la prueba brinden una aproximación apropiada de los niveles de habilidad de los sujetos en el constructo medido por la prueba (AERA, APA, & NCME, 2014).

Generación de conclusiones

Una vez calificados los exámenes, se proce-

de a la generación de interpretaciones con respecto al análisis de las puntuaciones, y a determinar parámetros para las inferencias pretendidas. Si el objetivo es establecer niveles de dominio en las dimensiones de la prueba, se debe recurrir a un proceso de establecimiento de estándares con los ítems seleccionados (*standard setting*). Si la finalidad es comparar el rendimiento de los sujetos contra el resto de la población, se puede recurrir al análisis de los percentiles de las puntuaciones obtenidas. Por último, se debe entregar un informe de calificación a los sujetos, indicándoles el significado de estas, ya que una nota por sí misma no dice nada con respecto a su desempeño en la prueba y puede crear nociones erróneas a los examinados, dado que culturalmente algunas notas se asocian a un buen rendimiento mientras que otras a un bajo rendimiento.

Dadas las etapas teóricas para la construcción de una prueba educativa, a continuación se presenta la descripción de cada una de ellas en la elaboración de la Prueba de Habilidades Cuantitativas en el caso específico del test del 2018.

Etapas de la construcción de la PHC previas a las aplicaciones regulares

En el 2003, las autoridades de la UCR decidieron que era necesario crear una prueba de ingreso a las carreras cuyas mallas curriculares tenían varios cursos de matemática. La prueba que se construyó para este propósito fue la PHC, la cual se aplica regularmente todos los años, desde el 2015. La construcción de los formularios de cada año se basa en las etapas iniciales de construcción de pruebas (desde el plan general hasta el pilotaje), que se desarrollaron durante varios años de investigación. A continuación, se describen los elementos principales de cada una de esas etapas.

Plan general

En esta etapa se determinó que el constructo razonamiento cuantitativo (RC) es lo que se quiere medir con la PHC; mientras que el uso (*para qué*) establecido fue brindar un criterio para la selección de los nuevos estudiantes que quisieran ingresar a carreras que requerían del uso de la matemática. Cabe mencionar que el RC se define como “la habilidad para analizar información cuantitativa y determinar cuáles destrezas y procedimientos pueden ser aplicados para obtener la solución de un problema particular” (Dwyer, Gallagher, Levin, & Morley, 2003, p. 1). Esta definición indica que el RC está conformado por los procesos de razonamiento con contenidos matemáticos que se deben realizar para llegar a la solución de un problema específico. En consecuencia, para la evaluación del RC se debe garantizar que los contenidos sobre los que se desarrollan las tareas de RC sean conocidos por toda la población.

En línea con la definición del RC, se puede decir que este no es equivalente al conocimiento matemático, ya que en el primero no es importante qué cantidad de conocimiento matemático tenga el sujeto, sino que este pueda determinar cómo utilizar el conocimiento que domina en situaciones particulares. Entre los componentes más importantes que demanda el RC de acuerdo con Niss y Højgaard (2011) están:

- el pensamiento matemático: consiste en utilizar las propiedades de los objetos matemáticos para llegar a una conclusión;
- el abordaje de problemas: consiste en diseñar una estrategia para utilizar un concepto matemático en la resolución de un problema;
- el razonamiento matemático: consiste en determinar la validez de proposiciones matemáticas; y
- la representación: consiste en diseñar o in-

interpretar una representación necesaria para llegar a la solución de un problema.

Con respecto al uso establecido para la PHC, se sostiene que el RC es necesario para desenvolverse exitosamente en las profesiones que requieren de la matemática en su quehacer (Mayes, 2019; Ryan & Gass, 2017) dado que los especialistas en estas áreas deben determinar cómo utilizar la matemática para resolver un problema determinado. Contrario a los ejercicios matemáticos clásicos de las carreras universitarias, en muchas tareas laborales no se indica qué algoritmo matemático se debe emplear, sino que el profesional debe construir una estrategia eficiente y pertinente que permita el éxito en la tarea. Con este fin, los especialistas de diversas áreas deben atender interrogantes utilizando gran variedad de información cuantitativa; esto los obliga a tomar decisiones sobre cuáles estadísticos brindan la información más adecuada a la pregunta atendida. En otros casos, se debe comparar fenómenos modelados por medio de expresiones algebraicas, lo cual demanda que el profesional distinga cuál es la estrategia que le permite realizar la comparación solicitada. Es por esto que se considera que los estudiantes de los programas de estudio que utilizan la matemática en su campo laboral deben poseer un nivel aceptable de RC (Rojas, Mora, & Ordóñez, 2018).

Ahora bien, dado que con la PHC se pretende discriminar a los sujetos con habilidad aceptable de aquellos con habilidades menores, se estableció que se necesitaba generar un punto de corte que separara a estas dos poblaciones, el cual puede ser representado en una escala de habilidad normal estándar como el valor ($\theta = 0$). Esto implicó que se utilizara un modelo de interpretación de puntajes con base a criterios y que se utilizara a la TRI como modelo de medición, ya que esta teoría permite maximizar la precisión en un nivel de habilidad específico (Martínez-Arias et al., 2006;

Muñiz, 1997).

Definición del contenido

Como se mencionó, los componentes del constructo RC establecidos en los ítems de la prueba son: el pensamiento matemático, el abordaje de problemas, el razonamiento matemático y la representación (Niss & Højgaard, 2011). Por otro lado, los conocimientos base que se utilizan en la prueba se estructuran de acuerdo con las áreas temáticas de la educación primaria y secundaria inicial de Costa Rica: análisis de datos, aritmética, álgebra y geometría. Solo se consideran contenidos de la secundaria inicial, dado que se asume que estos son los dominados por los estudiantes aspirantes a ingresar en las carreras mencionadas. La inclusión de contenidos poco dominados puede generar una fuente de varianza irrelevante al constructo pretendido.

Especificaciones del test

La tabla de especificaciones (Tabla 1) de la prueba se planteó como una matriz de procesos, según los componentes y áreas de contenido, con una distribución homogénea. Es importante mencionar que aunque se postulan distintos procesos, estos no se conciben como independientes entre sí (Niss & Højgaard, 2011), sino que el nombre del proceso refleja aquel que se considera más demandante en la solución del ítem. Por lo general los cuatro componentes se emplean para la resolución del reactivo en algún grado. Por lo tanto, se considera que el modelo factorial que mejor representa a los ítems es el unidimensional.

Por otro lado, como la población meta de la prueba era numerosa y el presupuesto limitado, se decidió recurrir a una prueba escrita de selección

única, con una hoja adicional en la que se consignen las respuestas. Es importante destacar que aunque la información brindada por los reactivos de selección única es valiosa, las preguntas de respuesta abierta podrían brindar mayor información de los procesos realizados por los sujetos, ya que se podrían evaluar los pasos de razonamiento efectuados para llegar a la respuesta o resolución del ítem.

Por otra parte, se estableció que los ítems debían tener una dificultad TRI promedio igual a 0, pues se buscaba maximizar la precisión en el nivel de habilidad ($\theta = 0$). Finalmente, por las características de los ítems, se determinó que 40 reactivos era una cantidad suficiente para alcanzar una precisión adecuada y que la resolución de una prueba de este tipo tomaría alrededor de dos horas. Cabe resaltar que se consideró que el cansancio asociado a resolver más ítems podría aumentar las fuentes de error de medición.

Tabla 1

Tabla de especificaciones pretendida para la PHC 2018.

Procesos	Área de contenido			
	Análisis de datos	Aritmética	Álgebra	Geometría
Pensamiento matemático	2 a 3 ítems	2 a 3 ítems	2 a 3 ítems	2 a 3 ítems
Representación	2 a 3 ítems	2 a 3 ítems	2 a 3 ítems	2 a 3 ítems
Abordaje de problemas	2 a 3 ítems	2 a 3 ítems	2 a 3 ítems	2 a 3 ítems
Razonamiento matemático	2 a 3 ítems	2 a 3 ítems	2 a 3 ítems	2 a 3 ítems

constructores entregaron los reactivos y estos fueron juzgados individualmente por el resto de los constructores y un juez adicional. En el juzgamiento se evaluó si los ítems realmente se ajustaban a los componentes de la tabla de especificaciones establecida y se analizó si tenían alguna de las fuentes de varianza irrelevante al constructo mencionadas previamente; luego, se procedió a un juzgamiento grupal para determinar si los ítems construidos podían ser considerados para el ensamblaje de la PHC.

Construcción de los ítems

El perfil para las personas encargadas de construir los ítems incluyó formación en enseñanza de la matemática y conocimientos de medición. En las capacitaciones de los constructores se presentaron el objeto de medida (RC) y sus componentes. Asimismo, se proporcionaron ejemplos de ítems en cada una de las categorías establecidas y se señalaron los elementos que agregaban varianza irrelevante al constructo: uso de algoritmos extensos, preguntas semejantes a las de contenido matemático, uso de tecnicismos y alta demanda de comprensión de lectura. Luego, a cada constructor se le solicitó que presentara un ítem de cada una de las casillas de la tabla de especificaciones.

Dos meses después de la capacitación, los

Pilotaje

Posteriormente, los ítems seleccionados fueron ensamblados en formularios con las características establecidas en la tabla de especificaciones y se aplicaron en muestras de estudiantes universitarios de primer año. Estas aplicaciones piloto permitieron desarrollar un banco de ítems de aproximadamente 80 reactivos con propiedades psicométricas adecuadas, según el modelo de TRI de dos parámetros. Estas propiedades se

rán explicadas en la etapa de calificación de los ítems. Las puntuaciones de estas aplicaciones brindaron evidencias de validez basadas en la relación con otras variables (específicamente, con el rendimiento académico, Bolaños-Barquero & Rojas-Torres, 2013; Rojas et al., 2018; Rojas-Torres, 2014) y de estructura interna (Rojas, 2013).

Elaboración del formulario de la PHC 2018

La construcción del formulario final utilizado en el 2018 se realizó con 36 de los ítems incluidos en el banco y 4 ítems nuevos. El pilotaje de estos ítems nuevos se llevó a cabo en la aplicación real de la prueba. Lo anterior implicó que estos ítems no fueran tomados en cuenta en la calificación. Se decidió bajar la cantidad de ítems calificables para asegurar que los nuevos ítems tuvieran estadísticas basadas en la situación real de evaluación en vez de basarse en una aplicación piloto.

Ensamblaje de la prueba

La selección final de los ítems tuvo la distribución que se presenta en la Tabla 2. Entre los criterios utilizados para seleccionar a los ítems de la tabla de especificaciones se consideró que no debían existir ítems con procesos de resolución muy semejantes entre sí, ya que esto podría pe-

nalizar doblemente al sujeto que no logró determinar el proceso de resolución óptimo. También se descartaron los ítems cuyas dificultades TRI, registradas en las aplicaciones piloto, se alejaron de la dificultad promedio pretendida ($\theta = 0$). Para este formulario la dificultad TRI promedio fue de ($\theta = .123$).

Para la administración de la PHC se proporcionaron varias sedes de aplicación en distintos edificios de entidades educativas ubicadas en zonas estratégicas de Costa Rica. Cada una de estas sedes tuvo entre 2 y 11 aulas, según la cantidad de personas inscritas en cada sede. Además, se coordinó con las autoridades de las entidades para que durante la administración de la prueba no hubiese personas ajenas a la aplicación cerca de las aulas y para que las aulas tuviesen iluminación y escritorios en buen estado.

Para cada sede se conformó un equipo de aplicación compuesto por un coordinador general, aplicadores para cada aula y un asistente del coordinador. Cada coordinador fue capacitado por el equipo desarrollador de la PHC y, luego, los coordinadores de sede capacitaron a sus equipos. En la capacitación se indicaron las pautas que se debían seguir para la administración de la prueba, por ejemplo: la forma de ordenar los muebles (pupitres) del aula, el protocolo para permitir el ingreso de los examinados a las aulas, las consultas que se podían responder y la vigilancia del grupo.

El día de la aplicación de la prueba, se leyó a los estudiantes unas instrucciones generales de

Tabla 2

Tabla de especificaciones final de la PHC 2018.

Procesos	Área de contenido			
	Análisis de datos	Aritmética	Álgebra	Geometría
Pensamiento matemático	3 ítems	2 ítems	3 ítems	2 ítems
Representación	2 ítems	3 ítems	3 ítems	3 ítems
Abordaje de problemas	3 ítems	2 ítems	2 ítems	3 ítems
Razonamiento matemático	2 ítems	3 ítems	2 ítems	2 ítems

cómo resolverla. Algunas instrucciones se relacionaron con: los materiales que podían utilizar, el tiempo asignado para la resolución y la forma de rellenar la hoja para respuestas. El total de personas que tomó la prueba fue de 2387 (1112 hombres y 1275 mujeres; 787 estudiantes de colegio privado y 1600 de público). En su mayoría eran estudiantes de secundaria que deseaban ingresar a una de las carreras de la UCR que utilizaba la PHC para la selección de los nuevos estudiantes.

Calificación de los ítems

Dado el carácter empírico de esta etapa, se utilizará una estructura clásica de análisis estadístico (procedimiento-resultados) para explicar su implementación. Además, en esta etapa se presentará el método de calificación utilizado para procesos de investigación, el cual no ha sido implementado aún. Actualmente, se reportan las calificaciones con el número de ítems correctos.

Procedimiento. Primeramente, se analizó la calidad global del instrumento. Para esto se estimó el índice de confiabilidad de constructos de modelos de ecuaciones estructurales, el cual se considera satisfactorio si es superior a .70 (Cea-D'Ancona, 2002). Además, se evaluó la hipótesis de la unidimensionalidad mediante un análisis factorial confirmatorio (AFC), esta hipótesis representa la estructura interna teorizada para la prueba. En el AFC se consideró que un buen ajuste era alcanzado si la raíz del error cuadrático medio de aproximación (RMSEA) era menor que .06, el índice de Tucker-Lewis (TLI) era mayor que .95, el índice de ajuste comparativo (CFI) era mayor que .95 y las cargas factoriales de los ítems eran superiores a .30 (Hu & Bentler, 1999; Cea-D'Ancona, 2002).

Por otro lado, para evaluar la calidad de los ítems se estimó el modelo de TRI de dos pará-

metros. Se valoró que los ítems no tuvieran dificultades extremas (menores que -3 o mayores que 3). También, se analizó que su discriminación fuera mínimamente aceptable (mayor a .35) y que brindaran información en el punto de interés ($\theta = 0$; función de información mayor que .10; Martínez-Arias et al., 2006).

Luego, con los ítems que cumplieron todos los criterios, se estimó la habilidad de los sujetos. Seguidamente, se analizó el error estándar de la estimación para verificar si este era bajo en el punto que se pretendía establecer como punto de corte ($\theta = 0$). Finalmente, se concluyó cuántas personas estuvieron en los grupos de habilidad que se requería discriminar ($\theta > 0$ y $\theta < 0$).

Análisis estadístico. Todos los análisis estadísticos se realizaron en la plataforma de programación estadística R, en su versión 3.3.2 (R Core Team, 2016). Los paquetes utilizados fueron *lavaan* (Rosseel, 2012), para la estimación de AFC, y *mirt* (Chalmers, 2012), para la estimación del modelo TRI.

Resultados. El índice de confiabilidad de constructos fue de .91, por lo cual se concluye que la prueba muestra evidencias de confiabilidad. Esta evidencia se basa en que los ítems son consistentes en la evaluación del constructo, es decir, las asociaciones de las puntuaciones de los ítems entre sí son semejantes. El indicador de consistencia interna usual de la TCT es el alfa de Cronbach, el cual fue de .85.

La estimación del modelo unidimensional del AFC se ajustó aceptablemente a los ítems de banco utilizados (CFI = .992, TLI = .992 y RMSEA = .014). No obstante, hubo 3 ítems (28, 29 y 35) que presentaron cargas factoriales inferiores a .30. El ajuste de este modelo brindó una *evidencia de validez basada en la estructura interna*.

Posteriormente, se analizaron las propieda-

des de los ítems según el modelo de TRI de dos parámetros. Se obtuvo que los mismos tres ítems que presentaron cargas factoriales bajas presentaban problemas en la discriminación (lo cual es esperable por la analogía de la carga factorial de un modelo unidimensional con la discriminación de la TRI de dos parámetros) y en la información en $\theta = 0$ (no solo fueron menores a .10, sino que fueron menores a .03). Hubo otros dos ítems (9 y 13) que presentaron valores de información en $\theta = 0$ ligeramente menores que el umbral establecido (.09 y .08), pero cumplieron el resto de los criterios establecidos. Por tanto, se decidió eliminar únicamente a los primeros tres ítems para la estimación de la habilidad.

Finalmente, la *calificación* de los sujetos correspondería a la habilidad de los sujetos a partir de los ítems seleccionados. Se obtuvo que el 44.0% de la población (1311 sujetos) tenía una habilidad mayor o igual a $\theta = 0$. Por otro lado, se recurrió a una de las ventajas de la TRI: la estimación del error estándar de la habilidad, lo cual permite la generación de intervalos de confianza. Se realizó la prueba de hipótesis para la hipótesis nula de $\theta \geq 0$, esta se rechazó con una significancia del 5% para el 30.5% de la población (911 personas); es decir, que el 30.5% de los sujetos presentaron una habilidad inferior a 0, con un 95% de confianza. De forma análoga, se obtuvo que el 24.7% de la población (739 personas) tuvo una habilidad superior a 0, con una confianza del 95%.

Generación de conclusiones

A partir de los datos de la calificación, la generación de conclusiones con propósitos de investigación determinaría que con un 95% de confianza, el 30.5% de la población no tiene la habilidad de RC requerida y que el 24.7% sí po-

see la habilidad requerida, por tanto, los primeros no pueden concursar por aquellas carreras en la que se utiliza la PHC; mientras que los segundos sí pueden participar. Con respecto al 44.7% de la población restante, no se pueden generar conclusiones tan contundentes.

La generación de conclusiones implementada oficialmente por la mayoría de las carreras se basa en un umbral asociado al porcentaje de respuestas correctas. Se considera que los sujetos con notas menores a este umbral no poseen la habilidad requerida para cursar estas carreras. El umbral se estableció con el cuidado de que las personas con notas inferiores a este punto presentaran habilidades inferiores a 0. A los sujetos con notas debajo del umbral se les indicó que no cumplieran con el requisito para concursar en las carreras. La razón por la que no se usa la habilidad TRI como criterio oficial es que es difícil de informar a los sujetos, ya que dos sujetos con la misma cantidad de aciertos no tendrán necesariamente el mismo nivel de habilidad.

Discusión

Este artículo muestra que el proceso de construcción de una prueba educativa es complejo y demandante. Es un proceso que no se puede realizar de manera ligera e irreflexiva. No obstante, la masificación del uso de estos instrumentos de evaluación ha generado que muchas de las etapas sean ignoradas. Esto induce a que las inferencias y los usos de las puntuaciones de las pruebas carezcan de validez.

En el caso de la PHC, las etapas iniciales de plan general, definición del contenido y especificaciones del test *requirieron* un período de reflexión de aproximadamente dos años; en esta discusión estuvieron involucrados expertos en medición, representantes de las carreras intere-

sadas y autoridades universitarias. El período de tiempo puede parecer excesivo, pero cuando se considera el uso propuesto para las puntuaciones de la prueba este tiende a ser razonable. En el caso de las pruebas educativas utilizadas en los salones de clase, se requiere que el tiempo dedicado a estas etapas iniciales sea mucho menor, pero esto no es excusa para eliminar la reflexión sobre ¿qué se desea medir?, ¿para qué se desea medir?, y ¿cómo se debe medir? Estas reflexiones pueden lograr evitar errores de medición clásicos de evaluación educativa.

En las etapas de construcción de los ítems y ensamblaje de la prueba se debe procurar que el formulario a desarrollar responda a las preguntas planteadas previamente. Además, en estas etapas, junto con la administración de la prueba, se debe procurar que los examinados tengan las condiciones necesarias para que logren mostrar su verdadero nivel de habilidad. En el caso de la PHC se invierte un período de aproximadamente ocho meses para el diseño de un formulario, este ciclo comienza con la construcción de los nuevos ítems, luego, el juzgamiento de los ítems, el ensamblaje de la prueba y la revisión del contenido del formulario. Como ya se mencionó, para las pruebas de aula no se puede exigir estos periodos de tiempo, pero no pueden ser eximidas de las reflexiones sobre si los ítems utilizados y la prueba en su totalidad miden lo que se desea medir y si las condiciones de aplicación realmente posibilitan que el examinado muestre su verdadera habilidad.

Con respecto a la calificación se debe tener la flexibilidad de variar la puntuación en función de la calidad de los ítems, ya que los mecanismos rígidos de calificación pueden llevar a interpretaciones erróneas, lo cual repercute drásticamente en la generación de conclusiones, y por ende perjudica a los examinados. Al menos, se debe procurar brindar a los sujetos las conclusiones reales

relacionadas con su rendimiento, dado que muchas veces hay creencias generalizadas sobre los significados de las calificaciones que no corresponden con el planteamiento teórico de la prueba; por ejemplo, un buen rendimiento es superior a un 7 de la escala de 0 a 10. En el caso de la PHC, actualmente se está trabajando en crear un informe de resultados que sea lo suficientemente informativo.

La última etapa en la construcción de una prueba es una de las más fallidas, pocas veces se le dice a un examinado qué significa la puntuación que obtuvo. Esta omisión es un problema ético importante, dado que se puede afectar la autoestima de un sujeto por una interpretación errónea, la cual es esperable debido a la atmósfera cultural. Debido a esto, en los estándares de evaluación psicológica y educativa se enfatiza sobre la forma de entregar los informes de calificación (AERA, APA, & NCME, 2014). Ahora bien, la interpretación de los puntajes solo es posible si la construcción de la prueba fue realizada siguiendo detalladamente las etapas de construcción de la misma, por tanto, se hace evidente la importancia de los protocolos de construcción de pruebas.

Es importante mencionar que la construcción de una prueba es un proceso continuo. En el caso de la PHC, después de los resultados de la aplicación del 2018, se concluyó que es necesario desarrollar un ensamblaje con ítems cuyas dificultades fueran cercanas a 0, en vez de que su promedio fuera cercano a 0, dado que esto permitiría aumentar la discriminación en el punto deseado y de esta manera disminuir el porcentaje de sujetos que no se puede clasificar en habilidades mayores que 0 o menores que 0. El abandono del promedio de la dificultad TRI se debe a que los ítems maximizan su discriminación en el nivel de dificultad; el promedio no implica ítems con dificultades en el nivel deseado. Por otro lado, el equipo constructor de la PHC está trabajando en

una definición de procesos más específicos para ciertos ítems de RC, en vez de los mencionados en este artículo, que son más comunes a todos los ítems de la prueba (Jeannotte & Kieran, 2017).

Finalmente, es importante mencionar que este artículo presenta un planteamiento general de las etapas necesarias para la construcción de una prueba educativa y que utiliza como ejemplo una prueba escrita de selección única, con un modelo de interpretación de puntajes basado en criterios y con modelo de medición de la TRI de dos parámetros. Los detalles específicos para cada variante de una prueba educativa no pueden ser presentados en un artículo de revista, ya que dependerán de la finalidad de la medición. No obstante, las etapas presentadas en este artículo incluyen los elementos mínimos que debe considerar cualquier prueba educativa.

Referencias

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for Educational and Psychological Testing*. Washington D. C.: American Educational Research Association.
- Bolaños-Barquero, M., & Rojas-Torres, L. (2013). Comparación entre los promedios de la Prueba de Aptitud Académica y la Prueba de Habilidades Cuantitativas de los estudiantes de la universidad de Costa Rica. *Revista de Ciencias Sociales*, 142(IV), 101-115. doi: [10.15517/rcs.v0i142.14305](https://doi.org/10.15517/rcs.v0i142.14305)
- Castillo-Arredondo, S., & Cabrerizo-Diago, J. (2010). *Evaluación educativa de aprendizajes y competencias*. Madrid, España: Pearson Educación.
- Cea-D'Ancona, M. A. (2002). *Análisis multivariable. Teoría y práctica en la investigación social*. Madrid, España: Síntesis.
- Chalmers, R. P. (2012). Mirt: A multidimensional item response theory package for the R environment. *Journal Of Statistical Software*, 48(6), 1-29. doi: [10.18637/jss.v048.i06](https://doi.org/10.18637/jss.v048.i06)
- Downing, S. M. (2006). Twelve steps for effective test development. En S. M. Downing & T. M. Haladyna (Eds.), *Handbook of Test Development* (pp. 3-26). Londres, Inglaterra: Lawrence Erlbaum. doi: [10.4324/9780203874776.ch1](https://doi.org/10.4324/9780203874776.ch1)
- Dwyer, C. A., Gallagher, A., Levin, J., & Morley, M. E. (2003). What is quantitative reasoning? Defining the construct for assessment purposes. *ETS Research Report Series*, 2003(2), 1-48. doi: [10.1002/j.2333-8504.2003.tb01922.x](https://doi.org/10.1002/j.2333-8504.2003.tb01922.x)
- Embretson, S. (2017). An integrative framework for construct validity. En A. A. Rupp & J. P. Leighton (Eds.), *The Handbook of Cognition and Assessment, Frameworks, Methodologies and Applications* (pp. 102-123). Oxford, MS: Willey Blackwell. doi: [10.1002/9781118956588](https://doi.org/10.1002/9781118956588)
- Ferrara, S., Lai, E., Reilly, A., & Nichols, P. D. (2017). Principled approaches to assessment design, development and implementation. En A. A. Rupp & J. P. Leighton (Eds.), *The Handbook of Cognition and Assessment, Frameworks, Methodologies and Applications* (pp. 41-74). Oxford, MS: Willey Blackwell. doi: [10.1002/9781118956588](https://doi.org/10.1002/9781118956588)
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1-55. doi: [10.1080/10705519909540118](https://doi.org/10.1080/10705519909540118)
- Jeannotte, D., & Kieran, C. (2017). A conceptual model of mathematical reasoning for school mathematics. *Educational Studies in Mathematics*, 96(1), 1-16. doi: [10.1007/s10649-017-9761-8](https://doi.org/10.1007/s10649-017-9761-8)
- Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50(1), 1-73. doi: [10.1111/jedm.12000](https://doi.org/10.1111/jedm.12000)
- Martínez-Arias, M. R., Hernández-Lloreda, M. J., & Hernández-Lloreda, M. V. (2006). *Psicometría*. Madrid, España: Alianza.
- Mateo, J., & Martínez, F. (2008). *Medición y evaluación*

- educativa*. Madrid, España: La Muralla.
- Mayes, R. (2019). Quantitative reasoning and its rôle in interdisciplinarity. En B. Doig, J. Williams, D. Swanson, R. Borromeo-Ferri & P. Drake (Eds.), *Interdisciplinary Mathematics Education. The State of the Art and Beyond ICEM 13 Monographs* (pp. 113-133). Cham, Suiza: Springer. doi: [10.1007/978-3-030-11066-6_8](https://doi.org/10.1007/978-3-030-11066-6_8)
- Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher*, 18(2), 5-11. doi: [10.3102/0013189X018002005](https://doi.org/10.3102/0013189X018002005)
- Muñiz, J. (1997). *Introducción a la Teoría de Respuesta a los Ítems*. Madrid, España: Pirámide.
- Muñiz, J., & Fonseca-Pedrero, E. (2019). Diez pasos para la construcción de un test. *Psicothema*, 31(1), 7-16. Recuperado de <http://www.psicothema.com>
- Niss, M., & Højgaard, T. (2011). *Competencies and Mathematical Learning. Ideas and inspiration for the development of mathematics teaching and learning in Denmark*. Roskilde, Dinamarca: IMFUFA.
- R Core Team. (2016). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Recuperado de <https://www.R-project.org>
- Rojas, L. (2013). Predicción de la dificultad de la Prueba de Habilidades Cuantitativas de la Universidad de Costa Rica. *Revista Digital Matemática, Educación e Internet*, 13(1), 1-14. Recuperado de <https://tecdigital.tec.ac.cr/revistamatematica>
- Rojas-Torres, L. (2014). Predicción de la reprobación de cursos de matemática básicos en las carreras de Física, Meteorología, Matemática, Ciencias Actuariales y Farmacia. *Revista Electrónica EDUCARE*, 18(3), 3-15. doi: [10.15359/ree.18-3.1](https://doi.org/10.15359/ree.18-3.1)
- Rojas, L., Mora, M., & Ordóñez, G. (2018). Asociación del razonamiento cuantitativo con el rendimiento académico en cursos introductorios de matemática de carreras STEM. *Revista Digital Matemática, Educación e Internet*, 19(1), 1-13. Recuperado de <https://tecdigital.tec.ac.cr/revistamatematica>
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal for Statistical Software*, 48(2), 1-36. doi: [10.18637/jss.v048.i02](https://doi.org/10.18637/jss.v048.i02)
- Ryan, A. M., & Gass, S. E. (2017). Quantitative reasoning: Exploring troublesome thresholds'. *Discussions on University Science Teaching: Proceedings of the Western Conference on Science Education*, 1(1), 1-16. Recuperado de <https://ir.lib.uwo.ca/wcsedust>
- Tiana, A. (1996). La evaluación de los sistemas educativos. *Revista Iberoamericana de Educación*, 10, 37-61. Recuperado de <https://rieoei.org/RIE>
- Villareal-Galera, M. P., Alfaro-Rojas, L., & Brizuela-Rodríguez, A. (2015). *Construcción de pruebas estandarizadas en el ámbito de la medición educativa y psicológica. Serie Cuadernos Metodológicos del IIP*. San José, Costa Rica: Instituto de Investigaciones Psicológicas. Recuperado de <http://www.kerwa.ucr.ac.cr>

Construction and validation of an ADHD screening tool for preschool children: A study with Spanish population

Construcción y validación de una escala de cribado de TDAH para niños preescolares: Un estudio con población española

Manuel Rodríguez-Becerra ¹, Luz María Fernández-Mateos ¹, Cristina Jenaro * ²

1 - Facultad de Educación, Universidad Pontificia de Salamanca, España.

2 - INICO/Facultad de Psicología, Salamanca, España.

Introduction
Method
Results
Discussion
References

Recibido: 26/01/2019 Revisado: 29/03/2019 Aceptado: 04/04/2019

Abstract

Most cases of ADHD occur in children under five, and this situation calls for screening tools with high sensitivity and specificity for early age. This study presents the ADHD-3P, a new screening tool developed with Spanish population for preschool children. From an initial pool of 151 items, face validity analyses, followed by item analyses and analysis of the scale, were carried out. A 27-item scale ($\alpha = .95$) composed of three factors: *hyperactivity* ($\alpha = .92$), *inattention* ($\alpha = .92$), and *other symptoms* ($\alpha = .81$) resulted from the analyses. Discriminant power was high (89.7% for ADHD, and 88.3% for non-ADHD). Diagnostic performance was evaluated through Receiver-Operating Characteristic curves, and excellent sensitivity (92.86%) and specificity (89.86%) were obtained. In sum, the ADHD-3P is a promising tool to be used by parents, teachers, and pediatricians.

Keywords: *attention-deficit/hyperactivity disorder, screening, preschoolers, psychometrics, measure development*

Resumen

La mayoría de los casos de TDAH suceden en niños menores de cinco años, y esta situación requiere de herramientas de cribado con elevada sensibilidad y especificidad para edades tempranas. Se presenta la ADHD-3P, una nueva herramienta de cribado desarrollada con población española para niños de preescolar. A partir de 151 ítems, se realizaron análisis de la validez aparente, análisis de los ítems y de la escala. Estos resultaron en una escala de 27 ítems ($\alpha = .95$) compuesta por tres factores: *hiperactividad* ($\alpha = .92$), *inatención* ($\alpha = .92$) y *otros síntomas* ($\alpha = .81$). El poder discriminativo de la escala fue elevado (89.7% para niños con TDAH y 88.3% para niños sin TDAH). La capacidad diagnóstica fue evaluada mediante curvas ROC y mostró una excelente sensibilidad (92.86%) y especificidad (89.86%). En suma, la ADHD-3P es una herramienta prometedora para ser utilizada por padres, profesores y pediatras.

Palabras clave: *TDAH, cribado, preescolar, propiedades psicométricas, desarrollo de escala*

*Correspondence to: Cristina Jenaro. INICO/Facultad de Psicología. Universidad de Salamanca. Avda. de la Merced, 109-131. 37005-Salamanca. Tel. (+34) 923294685. Fax. (+34) 923294695. E-mail: crisje@usal.es

How to cite: Rodríguez-Becerra, M., Fernández-Mateos, L. M., & Jenaro, C. (2019). Construction and validation of an ADHD screening tool for preschool children: A study with Spanish population. *Revista Evaluar*, 19(2), 30-42. Recuperado de <https://revistas.unc.edu.ar/index.php/revaluar>

Introduction

Attention Deficit Hyperactivity Disorder (ADHD; [American Psychiatric Association, 2013](#)), or Attention and Activity Disorder ([World Health Organization, 2010](#)) is one of the most prevalent neurodevelopmental disorders in the child population. Both the DSM (Diagnostic and Statistical Manual of Mental Disorders; [American Psychiatric Association, 2013](#)) and the ICD (International Classification of Diseases; [World Health Organization, 2010, 2018](#)) manuals designate attention deficit, hyperactivity and impulsivity as core symptoms. An overall prevalence of between 8% and 12% is estimated ([American Psychiatric Association, 2013](#); [World Health Organization, 2010, 2018](#)) and there is a general consensus regarding the underestimation of the ADHD ([Rowland et al., 2001](#)). Regarding clinical presentations, there is some agreement that inattentiveness is more common, and that for every girl diagnosed with ADHD there are three boys ([Willkut, 2012](#)).

The impact of ADHD on children and their families is widely known, hence the importance of early detection ([Hurtig, Taanila, Moilanen, Nordström, & Ebeling, 2012](#); [Kooij et al., 2010](#); [Sikirica et al., 2015](#); [Wehmeier, Schacht, & Barkley, 2010](#)). But detection is not easy for several reasons: (1) the diversity of clinical presentations ([Costa-Dias et al., 2013](#); [Potter, Dunbar, Mazzulla, Hosford, & Newhouse, 2014](#)); (2) the onset of the disorder coincides with some age-related behaviors ([Reyes-Sandoval & Acuña, 2012](#); [Vaquerizo-Madrid, 2005](#)); (3) the symptoms overlap with other disorders, such as Oppositional Defiant Disorder or Disruptive Mood Dysregulation Disorder ([Mulraney et al., 2016](#)) and (4) some of the behavioral disorders overlap with other emotional disorders, such as anxiety-related disorders ([Overgaard, Aase, Torgersen, & Zeiner, 2016](#)). In

sum, ADHD, rather than being a homogeneous entity, is a group of conditions with potentially different etiological and risk factors and with different outcomes ([Costa-Dias et al., 2013](#); [Spencer, Biederman, & Mick, 2007](#)).

This situation calls for screening tools with high sensitivity and specificity for early age, as most cases of ADHD occur in children under five ([Vaquerizo-Madrid, 2005](#)). Existing scales, such as the ADHD Symptom Checklist-4 (ADHD-SC4), have shown moderate sensitivity ([Sprafkin, Gadow, & Nolan, 2001](#)). The Attention Problem Scale of the Child Behavior Checklist (CBCL-APS) has shown moderate diagnostic performance ([Lampert, Polanczyk, Tramontina, Mardini, & Rohde, 2004](#)), as has the Pediatric Attention Disorders Diagnostic Screener (PADDS; [Newman & Reddy, 2017](#)) and the Persian version of the Conners' Adult Attention-Deficit/Hyperactivity Disorder (ADHD) Screening Scale ([Davari-Ashtiani, Jazayeri, Arabgol, Razjouyan, & Khademi, 2014](#)). Moderate levels of discriminant power have been obtained with the Face Stimulus Assessment (FSA; [Betts, 2003](#); [Kim, Kim, & Seo, 2014](#)) and other novel tools such as the AAPS ([Brownlie, Lazare, & Beitchman, 2012](#)). Additionally, moderate agreement levels have been found between diagnoses obtained with different screening tools ([Posserud et al., 2014](#)).

While several scales for assessing ADHD in preschool populations through proxies or key informants are currently available, there are few studies that prove the diagnostic capability of the screening tools in early childhood. Also, existing studies show a relatively low diagnostic capability ([Holmberg, Sundelin, & Hjern, 2013](#); [Lampert et al., 2004](#)), which results in false positives ([Sayal, Letch, & El Abd, 2008](#)).

Additional limitations are related to poor agreement in the assessment of symptoms ([Levin-Decanini, Connolly, Simpson, Suarez, &](#)

Jacob, 2013; Re & Cornoldi, 2009) and a lack of consensus between parents' and teachers' assessments (Deb, Dhaliwal, & Roy, 2008; Efst-ratopoulou, Simons, & Janssen, 2013; Power, Costigan, Leff, Eiraldi, & Landau, 2001; Raiker et al., 2017; Re & Cornoldi, 2009; Wolraich et al., 2004). The choice of screening strategy significantly affects how teachers report on ADHD symptoms at school. The halo effect of externalizing behaviors impacts on the correct identification of true cases of ADHD in the school setting (Kieling et al., 2014).

Another limitation of existing studies relates to the need to take cultural factors into account, as these even affect the perception of the severity of ADHD symptoms (CEAL-TDAH, 2009; Hillemeier, Foster, Heinrichs, & Heier, 2007; Norvilitis, Ingersoll, Zhang, & Jia, 2008). Hence the importance of instruments adapted to culture (Cornelio-Nieto, Borbolla-Sala, & García-Valdovinos, 2010). In sum, policies, research, and applied decisions depend on measurement instruments and the quality of decisions depends on the quality of the instruments (Danner et al., 2016; Ziegler, 2014).

Thus, in order to help overcome current shortcomings, this study developed the ADHD-3P scale. The name refers to the three informants who best know the child: Parents, Professors (teachers) and Pediatricians. This paper presents the process used to investigate the psychometric properties of the scale. First, we examined the adequacy of content; second, we examined the face validity of the scale. Next, we assessed its reliability and validity, which is understood as the degree to which all the accumulated evidence supports the intended interpretation of test scores for the intended purposes (AERA, APA, & NCME, 1999, as cited in Goodwin & Leech, 2003). It includes evidence from experts and judges' reviews on the test content, evidence on internal structure

by means of exploratory factor analysis and evidence on relations to other variables, specifically construct validity, and convergent and discriminant validity (Goodwin & Leech, 2003). Lastly, we examined the sensitivity and specificity of the scale for early identification of the first clinical manifestations compatible with ADHD and the cutoffs that distinguish between the clinical and nonclinical populations.

Method

Participants

This study required the participation of four different groups for the next steps (see Table 1). The first step required expert raters for face validity. A total of 16 experts, consisting of parents, professors, and pediatricians were consulted. The second step included a large sample of parents (N = 644) of children with (n = 281) and without ADHD (n = 363), to perform item analyses as well as Cronbach's alpha reliability tests and exploratory factor analysis. The third step utilized a sample of 79 informants who filled out the measure twice, for test-retest analyses. The fourth step utilized a sample of 54 parents of children with (n = 35) and without ADHD (n = 19) to determine the diagnostic accuracy of the developed measure. In all cases, confidentiality and voluntariness were guaranteed and the provisions of the participating agencies ethics' committees have been followed.

Measures

In addition to the ADHD-3P scale, whose development and characteristics are detailed in this research, the SNAP-IV Teacher and Parent Rating Scale (18 items; Swanson et al., 2001) was

used for the analysis of convergent validity and comparison of the diagnostic performance of both scales. This test is a hetero-report, completed by parents and teachers of children between 3 and 17 years to assess the presence and severity of behaviors consistent with ADHD. The 18 items (nine questions in the *attention* subscale and nine for *hyperactivity / impulsivity*) are scored on a Likert scale from 0 (*none*) to 3 (*very much*) points. Several studies support the sensitivity and specificity of the SNAP-IV (Alda & Serrano-Troncoso, 2013; Berrocal et al., 2011; Bussing et al., 2008) and so, the demonstrated quality of that measure served as the basis for its use in this study.

Procedure

The instrument was developed using the procedure summarized in Table 1. The preliminary step consisted in the development of an initial set of 151 items commonly used for assessing ADHD in children. The authors of the current

study were responsible for their initial selection from existing measures in different languages. Clinical criteria and clinical experience with the targeted population were also the basis for the development of some of the items. The number of initial items was considered reasonable given the expected dimensions and the final configuration of the screening tool (Netemeyer, Bearden, & Sharma, 2003).

Results

The ratings obtained by the 16 judges (n = 8 teachers, n = 5 parents, and n = 3 pediatricians) led to the retention of 45 (29.8%) of the original 151 items, which were found relevant, intense and clear enough for preschool children. Relevance and intensity were defined as an average score higher than 2.5 in a 1 to 4 scale. The clarity of the items was tested with the most common reliability coefficients for multiple coders assessing nominal data: the average pairwise percent agreement, Fleiss' kappa, average pairwise Co-

Table 1

Summary of the procedure (steps and tasks) followed in the development and validation of the ADHD-3P scale.

Steps and Tasks

Step 1: (Face validity) Consultation of 16 expert judges and ratings on:

1. Category
2. Intensity
3. Relevance

Step 2: Construction and application of a (45-item) preliminary version of the scale

1. Item Analyses (discriminative power, internal consistency)

Step 3: Application and analysis of the scale (27-item)

1. Factor analysis
 2. Internal consistency
 3. Agreement between raters
 4. Stability (test-retest)
 5. Construct validity
 6. Convergent validity: correlations with the SNAP IV
 7. Discriminant validity
 8. Diagnostic capability (obtaining ROC curves) and testing with ADHD preschool children
-

hen's kappa, and the Krippendorff's alpha (Free-lon, 2010). Next, the items analyses required to calculate the discriminative power of the items by means of *t*-test for independent groups (i.e. with or without ADHD). All items reached a confidence level of 1 per 1000 ($p < .001$). Then, we calculated the internal consistency coefficients of the items: Items with low corrected homogeneity (defined as values below .20) were excluded from further analysis resulting in 27 items.

Step 3, analysis of the scale (see Table 1), involved eight tasks. First, exploratory factor analysis was performed with the remaining 27 items. Principal component analysis with oblimin

rotation was performed, as a correlation between the extracted factors was expected, provided that the samples were big enough (KMO-test greater than .5; Field, 2000; Tabachnick & Fidell, 2007). Analyses led to the identification of three factors (loadings between .35 and .92). Thus, the final scale comprised of 27 items grouped into three factors. Table 2 shows the eigenvalues and percentages of explained variance. Analysis of subscales internal consistency was then performed and the findings indicated high reliability for *hyperactivity* ($\alpha = .92$), *inattention* ($\alpha = .92$) and *other symptoms* -impulsivity, irritability- ($\alpha = .81$), as well as for the scale total scores ($\alpha = .95$).

Table 2

Eigenvalues and explained variance (three-factor solution with varimax rotation).

Component	Total	% variance	% accumulated
1 Hyperactivity (n = 12)	10.38	38.43	38.43
2 Inattention (n = 9)	3.74	13.85	52.28
3 Other symptoms (n = 6)	1.71	6.32	58.60

Note. Method of extraction: principal components analysis. Oblimin rotation.

Task 3 (see Table 1) was to assess the agreement between raters. The analyses resulted in significant medium-high correlations between the ratings of both parents of the children with ADHD ($r = .52$ for inattention, $r = .80$ for hyperactivity, and $r = .80$ for other symptoms). The fourth task, determining the stability of the assessments, was calculated using the test-retest correlation between ratings of the informants. Values were higher than .98 for the different dimensions, which guarantees high stability.

The fifth task (Table 1), construct validity of the scale, required calculating correlations between the dimensions of the scale and age at diagnosis. As expected, the correlation between *hyperactivity* and age at diagnosis was significant and negative ($r = -.20$; $N = 277$; $p = .001$). The association was also significant with *inattention* (r

$= -.15$; $N = 277$; $p = .012$), and with *other symptoms* ($r = -.21$; $N = 277$; $p < .001$).

The sixth task required determining the convergent validity of the ADHD-3P with the SNAP IV scale through the correlations between scores on the factors and total scale. Correlations were high for *hyperactivity* ($r = .80$; $p < .01$), *inattention* ($r = .83$; $p < .01$) and the total scale ($r = .83$; $p < .01$). The size of the correlations further suggests that while similar content is being assessed, they are not identical, thus the scale under construction adds relevant and new content.

The seventh task (Table 1) was the analysis of the discriminant validity, taking the dimensions of the scale as potentially predictive variables and group membership (i.e. with or without ADHD) as a predicted variable. This led to the obtainment of a discriminant function with an eigenvalue = 1.561 and canonical correlation = .78,

which accounts for 100% of the variance (Wilks' = .39; $p = .0001$). Intragroup correlations of each variable with the canonical function were: .95 (inattention), .77 (hyperactivity) and .64 (other symptoms). In total, the function correctly classified for 88.5% of cases and showed high ability to classify participants with ADHD (i.e. sensitivity; 89.7% of cases correctly classified, CI: 85.52% to 92.98%), and those without ADHD (i.e. specificity; 88.3% of cases classified correctly, 95% CI = 84.56% to 91.46%).

The eighth task was the comparison of the diagnostic performance of the ADHD-3P against the SNAP-IV, using a sample of 54 cases of which 35 were clinical (i.e. with ADHD) and 19 were non-clinical (i.e. without ADHD). The ADHD-3P values of the areas under the curve (AUC) revealed the existence of great scale capacity to discriminate between clinical and non-clinical in *hyperactivity* and *inattention* (Table 3). Confidence intervals were equally satisfactory for the dimensions of *hyperactivity* and *inattention*, and slightly lower for *other symptoms*. The data were higher than those obtained with the SNAP which only include the two main dimensions of the disorder.

After obtaining the cutoff scores for each of the dimensions that maximize the sensitivity and specificity, we proceeded to test the diagnostic performance of the scale with preschool children ($n = 69$ nonclinical, and $n = 28$ clinical cases), through Receiver-Operating Characteristic (ROC) curves. Regarding *hyperactivity*, a cutoff score of 15.10 correctly identified the 26 clinical cases and 53 nonclinical cases. On the other hand, 16 false positives and 2 false negatives were identified. The sensitivity was 92.86% and the specificity was 76.81%. As for *inattention*, a cutoff score of 13 correctly identified 24 clinical cases and 65 nonclinical cases. It also identified 4 false positives and 4 false negatives. The sensitivity was 85.71% and the specificity was 94.20%. Regarding the *other symptoms* scale, a cutoff score of 7.50 correctly identified 27 clinical cases and 60 nonclinical cases. It also identified 9 false positives and 1 false negative. The sensitivity was 96.43% and specificity was 86.96%. Finally, considering the ADHD-3P scale overall, a cutoff score of 37.50 correctly identified 26 clinical and 62 nonclinical cases. It also identified 7 false positives and 2 false negatives. The sensitivity was 92.86% (95% CI = 76.50% to 99.12%) and speci-

Table 3

Area Under the Curve (AUC) of the variables Hyperactivity, Inattention and Other symptoms.

Variables	AUC	SE	p	95% asymptotic Confidence limits	
				Lower Limit	Upper Limit
ADHD-3P					
Hyperactivity	.98	.01	< .001	.96	1.01
Inattention	.98	.02	< .001	.95	1.02
Other symptoms	.86	.06	< .001	.75	0.97
Total	.97	.03	< .001	.93	1.02
SNAP					
Hyperactivity	.97	.02	< .001	.93	1.01
Inattention	.97	.03	< .001	.92	1.03
Total	.99	.01	< .001	.97	1.01

ficity was 89.86% (95% CI = 80.21% to 95.82%). The positive likelihood ratio was 9.15 (95% CI = 4.50 to 18.61) and negative likelihood ratio was .08 (95% CI = .02 to .31). All results were fully satisfactory.

Discussion

In this research, we have presented the process utilized to build and validate a brief scale for screening children in early childhood with possible ADHD. The developed scale (see Appendix 1 and 2 for English and Spanish versions, respectively) has shown very high sensitivity for the ADHD diagnosis in Spanish preschool children. This sensitivity is particularly high for the diagnosis of hyperactive symptoms as well as other symptoms. It also provides better results than similar studies with other screening tools (Abdekhodaie, Tabatabaei, & Gholizadeh, 2012; Alda & Serrano-Troncoso, 2013; Holmberg et al., 2013). Consequently, this measure contributes to the improvement of the early detection of possible clinical cases of ADHD at early ages, which is always challenging from a clinical point of view. An additional noteworthy fact is that the measure was developed by asking three different groups of informants: parents, teachers, and pediatricians. Each group contributed their visions of the disorder under consideration and their different experiences with it as well as the context (clinical, educational, family) when they need to face it for different purposes (diagnosis, treatment, support, etc.). These differences in purposes and roles offered a broader and more comprehensive view of the disorder and increased the content validity of the developed measure at the same time.

The high reliability obtained, understood as internal consistency, measurement stability and concordance between evaluators, supports great-

er measure accuracy than shown in other studies (Erhart, Döpfner, & Ravens-Sieberer, 2008; Kim et al., 2014). The analysis of content validity (by way of judges), construct validity (through factor analysis), convergent validity (with SNAP) and discrimination (using discriminant analysis and ROC curves) supports confidence in the adequacy of the scale for assessing relevant symptoms in Spanish children with ADHD, and improves on the results of other similar studies with other measures (Bussing et al., 2008; Sprafkin et al., 2001). The diagnostic performance of the ADHD-3P was also higher than that obtained with the SNAP, which only includes the two main dimensions of the disorder. The domain other symptoms includes items on impulsivity, which is acquiring increasing relevance in current factorial models on ADHD (Dumenci, McConaughy, & Achenbach, 2004).

However, this study is only a first step in the development and consolidation of screening tools for ADHD, and caution against using the ADHD-3P as the only tool for diagnostic purposes is advised. Currently, the predictive ability of the instrument is unknown. This will require the incorporation of children at risk without a diagnosis and subsequent evaluation of sensitivity and specificity. The continued use and dissemination of the ADHD-3P in Spanish and English speaking countries will allow us to answer this and other questions. The inclusion of the questionnaire in the appendix aims to serve as an incentive for use and dissemination in subsequent research efforts.

Yet, despite what we believe is a clear contribution in the field in question, we would like to put an emphasis on the importance of going beyond assessments and not focusing solely on the child. Effective interventions (Mikami et al., 2013) should include not only supports for the child but also for their parents and teachers. Hence the importance of targeting the “three Ps”:

parents, professors, and pediatricians involved in the process of assessment and intervention.

References

- Abdekhodaie, Z., Tabatabaei, S. M., & Gholizadeh, M. (2012). The investigation of ADHD prevalence in kindergarten children in northeast Iran and a determination of the criterion validity of Conners' Questionnaire via clinical interview. *Research in Developmental Disabilities, 33*(2), 357-361. doi: [10.1016/j.ridd.2011.10.006](https://doi.org/10.1016/j.ridd.2011.10.006)
- Alda, J. A., & Serrano-Troncoso, E. (2013). Trastorno por déficit de atención con hiperactividad: Concordancia entre la impresión clínica y el cuestionario de cribado SNAP-IV (Estudio CONCOR). *Actas Españolas de Psiquiatría, 41*(2), 76-83. Retrieved from <https://www.actaspsiquiatria.es>
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders. DSM-5* (5th ed.). Washington, DC: American Psychiatric Association. doi: [10.1176/appi.books.9780890425596](https://doi.org/10.1176/appi.books.9780890425596)
- Berrocal, M., Peskin, V., Weiss, N., Schuler, J., Monge, S., McGough, J. J., ... Mathews, C. A. (2011). Prevalencia y tamizaje del trastorno por déficit de atención con hiperactividad en Costa Rica. *Vertex, Revista de Psiquiatría, 22*(99), 337-342. Retrieved from <http://www.polemos.com.ar/vertex.php>
- Betts, D. J. (2003). Developing a projective drawing test: Experiences with the Face Stimulus Assessment (FSA). *Art Therapy: Journal of the American Art Therapy Association, 20*(2), 77-82. doi: [10.1080/07421656.2003.10129393](https://doi.org/10.1080/07421656.2003.10129393)
- Brownlie, E., Lazare, K., & Beitchman, J. (2012). Validating a self-report screen for ADHD in early adulthood using childhood parent and teacher ratings. *Journal of Attention Disorders, 16*(6), 467-477. doi: [10.1177/1087054711398902](https://doi.org/10.1177/1087054711398902)
- Bussing, R., Fernandez, M., Harwood, M., Wei Hou, Garvan, C. W., Eyberg, S. M., & Swanson, J. M. (2008). Parent and teacher SNAP-IV ratings of attention deficit hyperactivity disorder symptoms: Psychometric properties and normative ratings from a school district sample. *Assessment, 15*(3), 317-328. doi: [10.1177/1073191107313888](https://doi.org/10.1177/1073191107313888)
- Comité Internacional para el Desarrollo y Estudio del Cuestionario para Escolares y Adolescentes Latinoamericanos con Trastorno por Déficit de Atención con Hiperactividad (CEAL-TDAH). (2009). Cuestionario para Escolares y Adolescentes Latinoamericanos con Trastorno por Déficit de Atención con Hiperactividad (CEAL-TDAH). Construcción, descripción del instrumento y datos sociodemográficos y clínicos de la muestra. *Salud Mental, 32*(Supl. 1), 55-62. Retrieved from <http://revistasaludmental.mx>
- Cornelio-Nieto, J. O., Borbolla-Sala, M. E., & García-Valdivinos, S. (2010). Population-based diagnostic study to compare the DSM-IV and the Questionnaire for screening Latin American schoolchildren and adolescents for attention deficit hyperactivity disorder. *Revista de Neurología, 50*(3), 119-123. doi: [10.33588/m.50S03.2010012](https://doi.org/10.33588/m.50S03.2010012)
- Costa-Dias, T. G., Kieling, C., Graeff-Martins, A. S., Moriyama, T. S., Rohde, L. A., & Polanczyk, G. V. (2013). Developments and challenges in the diagnosis and treatment of ADHD. *Revista Brasileira de Psiquiatria, 35*(1), 40-50. doi: [10.1590/1516-4446-2013-S103](https://doi.org/10.1590/1516-4446-2013-S103)
- Danner, D., Blasius, J., Breyer, B., Eifler, S., Menold, N., Paulhus, D. L., ... Ziegler, M. (2016). Current challenges, new developments, and future directions in scale construction. *European Journal of Psychological Assessment, 32*(3), 175-180. doi: [10.1027/1015-5759/a000375](https://doi.org/10.1027/1015-5759/a000375)
- Davari-Ashtiani, R., Jazayeri, F., Arabgol, F., Razjouyan, K., & Khademi, M. (2014). Psychometric properties of Persian version of Conners' Adult Attention Deficit/Hyperactivity Disorder Rating Scale (screening form-self reporting). *Iranian Journal of Psychiatry and Clinical Psychology, 20*(3), 243-251.

- Deb, S., Dhaliwal, A. J., & Roy, M. (2008). The usefulness of Conners' Rating Scales-Revised in screening for attention deficit hyperactivity disorder in children with intellectual disabilities and borderline intelligence. *Journal of Intellectual Disability Research, 52*(11), 950-965. doi: [10.1111/j.1365-2788.2007.01035.x](https://doi.org/10.1111/j.1365-2788.2007.01035.x)
- Dumenci, L., McConaughy, S. H., & Achenbach, T. M. (2004). A hierarchical three-factor model of inattention hyperactivity-impulsivity derived from the attention problems syndrome of the teacher's report form. *School Psychology Review, 33*(2), 287-301. Retrieved from <https://www.nasponline.org/>
- Efstratopoulou, M., Simons, J., & Janssen, R. (2013). Concordance among physical educators', teachers', and parents' perceptions of attention problems in children. *Journal of Attention Disorders, 17*(5), 437-443. doi: [10.1177/1087054711431698](https://doi.org/10.1177/1087054711431698)
- Erhart, M., Döpfner, M., & Ravens-Sieberer, U. (2008). Psychometric properties of two ADHD questionnaires: Comparing the Conners' scale and the FBB-HKS in the general population of German children and adolescents - Results of the BELLA study. *European Child & Adolescent Psychiatry, 17*(1), 106-115. doi: [10.1007/s00787-008-1012-1](https://doi.org/10.1007/s00787-008-1012-1)
- Field, A. P. (2000). *Discovering statistics using SPSS for Windows*. London, England: Sage.
- Freelon, D. G. (2010). ReCal: Intercoder reliability calculation as a web service. *International Journal of Internet Science, 5*(1), 20-33. Retrieved from <http://www.ijis.net>
- Goodwin, L. D., & Leech, N. L. (2003). The meaning of validity in the new standards for educational and psychological testing: Implications for measurement courses. *Measurement and Evaluation in Counseling and Development, 36*(3), 181-191. doi: [10.1080/07481756.2003.11909741](https://doi.org/10.1080/07481756.2003.11909741)
- Hillemeier, M. M., Foster, E. M., Heinrichs, B., & Heier, B. (2007). Racial differences in parental reports of attention-deficit/hyperactivity disorder behaviors. *Journal of Developmental and Behavioral Pediatrics, 28*(5), 353-361. doi: [10.1097/DBP.0b013e31811ff8b8](https://doi.org/10.1097/DBP.0b013e31811ff8b8)
- Holmberg, K., Sundelin, C., & Hjern, A. (2013). Screening for attention-deficit/hyperactivity disorder (ADHD): Can high-risk children be identified in first grade? *Child: Care, Health and Development, 39*(2), 268-276. doi: [10.1111/j.1365-2214.2012.01382.x](https://doi.org/10.1111/j.1365-2214.2012.01382.x)
- Hurtig, T., Taanila, A., Moilanen, I., Nordström, T., & Ebeling, H. (2012). Suicidal and self-harm behaviour associated with adolescent attention deficit hyperactivity disorder-A study in the Northern Finland birth cohort 1986. *Nordic Journal of Psychiatry, 66*(5), 320-28. doi: [10.3109/08039488.2011.644806](https://doi.org/10.3109/08039488.2011.644806)
- Kieling, R. R., Kieling, C., Aguiar, A. P., Costa, A. C., Dorneles, B. V., & Rohde, L. A. (2014). Searching for the best approach to assess teachers' perception of inattention and hyperactivity problems at school. *European Child & Adolescent Psychiatry, 23*(6), 451-459. doi: [10.1007/s00787-013-0466-y](https://doi.org/10.1007/s00787-013-0466-y)
- Kim, J., Kim, G., & Seo, S. (2014). Validation of the FSA as screening tool for children with ADHD. *The Arts in Psychotherapy, 41*(4), 413-423. doi: [10.1016/j.aip.2014.04.006](https://doi.org/10.1016/j.aip.2014.04.006)
- Kooij, S. J. J., Bejerot, S., Blackwell, A., Caci, H., Casas-Brugué, M., Carpentier, P. J., ... Asherson, P. (2010). European consensus statement on diagnosis and treatment of adult ADHD: The European Network Adult ADHD. *BMC Psychiatry, 10*, 1-24. doi: [10.1186/1471-244x-10-67](https://doi.org/10.1186/1471-244x-10-67)
- Lampert, T. L., Polanczyk, G., Tramontina, S., Mardini, V., & Rohde, L. A. (2004). Diagnostic performance of the CBCL-Attention Problem Scale as a screening measure in a sample of Brazilian children with ADHD. *Journal of Attention Disorders, 8*(2), 63-71. doi: [10.1177/108705470400800204](https://doi.org/10.1177/108705470400800204)
- Levin-Decanini, T., Connolly, S. D., Simpson, D., Suarez, L., & Jacob, S. (2013). Comparison of behavioral profiles for anxiety-related comorbidities including ADHD and selective mutism in children. *Depression and Anxiety, 30*(9), 857-864. doi: [10.1002/da.22094](https://doi.org/10.1002/da.22094)
- Mikami, A. Y., Griggs, M. S., Lerner, M. D., Emeh, C. C., Reuland, M. M., Jack, A., & Anthony, M. R. (2013). A randomized trial of a classroom intervention to

- increase peers' social inclusion of children with attention-deficit/hyperactivity disorder. *Journal of Consulting and Clinical Psychology*, 81(1), 100-112. doi: [10.1037/a0029654](https://doi.org/10.1037/a0029654)
- Mulraney, M., Schilpzand, E. J., Hazell, P., Nicholson, J. M., Anderson, V., Efron, D. E., & Silk, T. J. (2016). Comorbidity and correlates of disruptive mood dysregulation disorder in 6-8-year-old children with ADHD. *European Child & Adolescent Psychiatry*, 25(3), 321-330. doi: [10.1007/s00787-015-0738-9](https://doi.org/10.1007/s00787-015-0738-9)
- Netemeyer, R. G., Bearden, W. O., & Sharma, S. (2003). *Scaling Procedures: Issues and Applications*. Thousand Oaks, CA: Sage. doi: [10.4135/9781412985772](https://doi.org/10.4135/9781412985772)
- Newman, E., & Reddy, L. A. (2017). Diagnostic utility of the Pediatric Attention Disorders Diagnostic Screener. *Journal of Attention Disorders*, 21(5), 372-380. doi: [10.1177/1087054714526431](https://doi.org/10.1177/1087054714526431)
- Norvilitis, J. M., Ingersoll, T., Zhang, J., & Jia, S. (2008). Self-reported symptoms of ADHD among college students in China and the United States. *Journal of Attention Disorders*, 11(5), 558-567. doi: [10.1177/1087054707308496](https://doi.org/10.1177/1087054707308496)
- Overgaard, K. R., Aase, H., Torgersen, S., & Zeiner, P. (2016). Co-occurrence of ADHD and anxiety in preschool children. *Journal of Attention Disorders*, 20(7), 573-580. doi: [10.1177/1087054712463063](https://doi.org/10.1177/1087054712463063)
- Posserud, M. B., Ullebø, A. K., Plessen, K. J., Stormark, K. M., Gillberg, C., & Lundervold, A. J. (2014). Influence of assessment instrument on ADHD diagnosis. *European Child and Adolescent Psychiatry*, 23(4), 197-205. doi: [10.1007/s00787-013-0442-6](https://doi.org/10.1007/s00787-013-0442-6)
- Potter, A. S., Dunbar, G., Mazzulla, E., Hosford, D., & Newhouse, P. A. (2014). AZD3480, a novel nicotinic receptor agonist, for the treatment of attention-deficit/hyperactivity disorder in adults. *Biological Psychiatry*, 75(3), 207-214. doi: [10.1016/j.biopsych.2013.06.002](https://doi.org/10.1016/j.biopsych.2013.06.002)
- Power, T. J., Costigan, T. E., Leff, S. S., Eiraldi, R. B., & Landau, S. (2001). Assessing ADHD across settings: Contributions of behavioral assessment to categorical decision making. *Journal of Clinical Child Psychology*, 30(3), 399-412. doi: [10.1207/s15374424jccp3003_11](https://doi.org/10.1207/s15374424jccp3003_11)
- Raiker, J. S., Freeman, A. J., Perez-Algorta, G., Frazier, T. W., Findling, R. L., & Youngstrom, E. A. (2017). Accuracy of Achenbach scales in the screening of attention-deficit/hyperactivity disorder in a community mental health clinic. *Journal of the American Academy of Child & Adolescent Psychiatry*, 56(5), 401-409. doi: [10.1016/j.jaac.2017.02.007](https://doi.org/10.1016/j.jaac.2017.02.007)
- Re, A. M., & Cornoldi, C. (2009). Two new rating scales for assessment of ADHD symptoms in Italian preschool children: A comparison between parent and teacher ratings. *Journal of Attention Disorders*, 12(6), 532-539. doi: [10.1177/1087054708323001](https://doi.org/10.1177/1087054708323001)
- Reyes-Sandoval, A. M., & Acuña, L. (2012). Juicios de maestros sobre las conductas características del trastorno por déficit de atención e hiperactividad. *Revista Latinoamericana de Psicología*, 44(3), 65-82. Retrieved from <https://editorial.konradlorenz.edu.co/rlp.html>
- Rowland, A. S., Umbach, D. M., Catoe, K. E., Stallone, L., Long, S., Rabiner, D., ... Sandler, D. P. (2001). Studying the epidemiology of attention-deficit hyperactivity disorder: Screening methods and pilot results. *The Canadian Journal of Psychiatry*, 46(10), 931-940. doi: [10.1177/070674370104601005](https://doi.org/10.1177/070674370104601005)
- Sayal, K., Letch, N., & El Abd, S. (2008). Evaluation of screening in children referred for an ADHD assessment. *Child and Adolescent Mental Health*, 13(1), 41-46. doi: [10.1111/j.1475-3588.2007.00463.x](https://doi.org/10.1111/j.1475-3588.2007.00463.x)
- Sikirica, V., Flood, E., Dietrich, C. N., Quintero, J., Harpin, V., Hodgkins, P., ... Haim-Erder, M. (2015). Unmet needs associated with attention deficit/hyperactivity disorder in eight European countries as reported by caregivers and adolescents: Results from qualitative research. *The Patient-Centered Outcomes Research*, 8(3), 269-281. doi: [10.1007/s40271-014-0083-y](https://doi.org/10.1007/s40271-014-0083-y)
- Spencer, T. J., Biederman, J., & Mick, E. (2007). Attention-deficit/hyperactivity disorder: Diagnosis, lifespan, comorbidities and neurobiology. *Journal of Pediatric Psychology*, 32(6), 631-642. doi: [10.1093/](https://doi.org/10.1093/)

[jpepsy/jism005](#)

- Sprafkin, J., Gadow, K. D., & Nolan, E. E. (2001). The utility of a DSM-IV-Referenced screening instrument for attention-deficit/hyperactivity disorder. *Journal of Emotional and Behavioral Disorders, 9*(3), 182-191. doi: [10.1177/106342660100900304](#)
- Swanson, J. M., Kraemer, H. C., Hinshaw, S. P., Arnold, L. E., Conners, C. K., Abikoff, H. B., ... Wu, M. (2001). Clinical relevance of the primary findings of the MTA: Success rates based on severity of ADHD and ODD symptoms at the end of treatment. *Journal of the American Academy of Child and Adolescent Psychiatry, 40*(2), 168-179. doi: [10.1097/00004583-200102000-00011](#)
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Upper Saddle River, NJ: Pearson Allyn & Bacon.
- Vaquero-Madrid, J. (2005). Hiperactividad en el niño preescolar: Descripción clínica. *Revista de Neurología, 40*(1), 25-32. doi: [10.33588/rn.40S01.2005037](#)
- Wehmeier, P. M., Schacht, A., & Barkley, R. A. (2010). Social and emotional impairment in children and adolescents with ADHD and the impact on quality of life. *Journal of Adolescent Health, 46*(3), 209-217. doi: [10.1016/j.jadohealth.2009.09.009](#)
- Willkut, E. G. (2012). The prevalence of DSM IV attention-deficit/hyperactivity disorder: A meta-analytic review. *Neurotherapeutics, 9*(3), 490-499. doi: [10.1007/s13311-012-0135-8](#)
- Wolraich, M. L., Lambert, E. W., Bickman, L., Simmons, T., Doffing, M. A., & Worley, K. A. (2004). Assessing the impact of parent and teacher agreement on diagnosing attention-deficit hyperactivity disorder. *Journal of Developmental and Behavioral Pediatrics, 25*(1), 41-47. doi: [10.1097/00004703-200402000-00007](#)
- World Health Organization. (2010). *International Classification of Diseases, 10th revision*. Geneva, Switzerland: World Health Organization.
- World Health Organization. (2018). *International Statistical Classification of Diseases and Related Health Problems, ICD-11*. Geneva, Switzerland: World Health Organization.
- Ziegler, M. (2014). Stop and state your intentions! Let's not forget the ABC of test construction. *European Journal of Psychological Assessment, 30*(4), 239-242.

Appendix 1.

ADHD-3P Scale for the screening of ADHD in early years of life (English version) (Rodríguez-Becerra, Fernández-Mateos, & Jenaro, 2019)

Instructions

Below there are a series of behaviors that your child may or may not manifest to different degrees. Please, respond to

each question by indicating the degree to which your child manifests the following behaviors using the following criteria or levels of intensity:

NOTHING if is not manifest = 0

LITTLE BIT if is not a relevant feature = 1

QUITE if it is a frequent behavior = 2

MUCH if is a typical feature for your child = 3

Please try to answer all questions. Thank you very much for your cooperation.

#	He or she:	N	L	Q	M	F
01.	Runs and jumps in inappropriate situations	0	1	2	3	H
02.	Is active	0	1	2	3	H
03.	Is impulsive and irritable (explosive episodes)	0	1	2	3	O
04.	Is restless, moving all the time	0	1	2	3	H
05.	It is prone to tantrums	0	1	2	3	O
06.	Talks excessively	0	1	2	3	H
07.	Fidgets with hands or feet or squirms in seat	0	1	2	3	H
08.	Is nervous	0	1	2	3	H
09.	Does not complete activities that he/she begins	0	1	2	3	I
10.	Does not grasp the meaning of "no"	0	1	2	3	O
11.	Is not invited to birthday parties or similar	0	1	2	3	O
12.	Does not seem to listen when spoken to	0	1	2	3	I
13.	Does not plan or develop a strategy for playing	0	1	2	3	I
14.	Does not pay attention to detail: is negligent	0	1	2	3	I
15.	Does not pay attention for long periods of time (10 minutes) or cannot concentrate	0	1	2	3	I
16.	Does not remember what he/she has done previously	0	1	2	3	I
17.	Is constantly running around	0	1	2	3	H
18.	Does not finish assignments	0	1	2	3	I
19.	Does not easily tolerate frustration	0	1	2	3	O
20.	Does not tolerate being still	0	1	2	3	H
21.	Has a difficult temperament	0	1	2	3	O
22.	Has difficulties playing quiet games	0	1	2	3	H
23.	Has difficulties organizing tasks	0	1	2	3	I
24.	Struggles to follow instructions and fails to finish tasks	0	1	2	3	I
25.	Shows difficulties waiting for his/her turn	0	1	2	3	H
26.	Has trouble with sitting down when required	0	1	2	3	H
27.	Touches everything although it is prohibited	0	1	2	3	H

Nota. F = Factor; H = Hyperactivity; I = Inattention; O = Other symptoms.

Appendix 2.

ADHD-3P Escala para el cribado de TDAH en los primeros años de vida (versión en castellano) (Rodríguez-Becerra, Fernández-Mateos, & Jenaro, 2019)

Instrucciones

A continuación se describen una serie de comportamientos que el niño/a puede o no manifestar en diferentes grados. Responda por favor a cada pregunta marcando con una cruz

el grado en que el niño/a manifiesta las conductas indicadas a continuación, respondiendo a los siguientes criterios o niveles de intensidad:

NADA o no lo manifiesta = 0

POCO o no es una característica relevante = 1

BASTANTE o es una conducta frecuente = 2

MUCHO o es una característica típica de su hijo/a = 3

Por favor, trate de responder a todas las preguntas. Muchas gracias por su colaboración.

N.º	El niño o niña:	N	P	B	M	F
01	Corre y salta en exceso en situaciones inapropiadas	0	1	2	3	H
02	Es excesivamente activo/a	0	1	2	3	H
03	Es impulsivo/a e irritable (episodios explosivos)	0	1	2	3	O
04	Es inquieto/a, no para de moverse	0	1	2	3	H
05	Es propenso/a a las rabietas	0	1	2	3	O
06	Habla excesivamente	0	1	2	3	H
07	Mueve en exceso las manos o los pies o se retuerce en su asiento	0	1	2	3	H
08	Es nervioso/a	0	1	2	3	H
09	No acaba las actividades que empieza	0	1	2	3	I
10	No entiende el significado de "NO"	0	1	2	3	O
11	No le invitan a las fiestas de cumpleaños o similar	0	1	2	3	O
12	No parece escuchar cuando se le habla	0	1	2	3	I
13	No planifica o elabora una estrategia para jugar	0	1	2	3	I
14	No presta atención a los detalles: es descuidado/a	0	1	2	3	I
15	No presta atención durante mucho rato (10 minutos) o no puede concentrarse	0	1	2	3	I
16	No recuerda lo que acaba de hacer	0	1	2	3	I
17	Se mueve constantemente	0	1	2	3	H
18	No termina lo que se le pide	0	1	2	3	I
19	No tolera fácilmente la frustración	0	1	2	3	O
20	No tolera la quietud	0	1	2	3	H
21	Muestra un temperamento difícil	0	1	2	3	O
22	Tiene dificultad para jugar a juegos tranquilos	0	1	2	3	H
23	Tiene dificultad para organizar las tareas	0	1	2	3	I
24	Tiene dificultad para seguir instrucciones y no logra terminar las tareas	0	1	2	3	I
25	Tiene dificultades para esperar su turno	0	1	2	3	H
26	Tiene problemas para sentarse cuando la situación lo requiere	0	1	2	3	H
27	Toca todo lo que ve aunque esté prohibido	0	1	2	3	H

Nota. F = Factor; H = Hiperactividad; I = Inatención; O = Otros síntomas.

Validación de una versión breve de la Escala de Satisfacción Sexual Subjetiva (ESSS-B) en Puerto Rico

Validation of a Short Version of the Subjective Sexual Satisfaction Scale (ESSS-B) in Puerto Rico

Juan Aníbal González-Rivera * ¹, Idania Hernández-Gato ²

1 - School of Behavioral and Brain Sciences, Ponce Health Sciences University, San Juan University Center, Puerto Rico.

2 - Universidad Carlos Albizu, San Juan Campus, Puerto Rico.

Recibido: 17/01/2019 Revisado: 25/02/2019 Aceptado: 10/03/2019

Introducción
Estudio 1
Método
Resultados
Estudio 2
Método
Resultados
Discusión
Referencias

Resumen

El objetivo principal de este proyecto fue desarrollar una versión breve de la Escala de Satisfacción Sexual Subjetiva. Para ello, se realizaron dos estudios independientes. El primero empleó una muestra de 200 participantes para analizar la confiabilidad y la estructura factorial de la escala con criterios estrictos de selección de ítems. En el segundo, se utilizó una muestra de 275 participantes, se efectuó un análisis factorial confirmatorio y se calculó la validez convergente y concurrente de la versión breve del instrumento (ESSS-B). Los resultados de los estudios mostraron propiedades psicométricas adecuadas en la escala. El índice de confiabilidad alfa de Cronbach de la escala fue .91. Los siete ítems de la ESSS-B cumplieron con los criterios de discriminación. Se realizaron análisis de validez convergente y concurrente, de los que se obtuvieron valores estadísticamente significativos. Estos resultados sugieren que la ESSS-B tiene el potencial para medir adecuadamente la satisfacción sexual de una forma global y subjetiva en adultos puertorriqueños.

Palabras clave: *satisfacción sexual, sexualidad, propiedades psicométricas, validación*

Summary

The objective of this study was to develop a short version of the Subjective Sexual Satisfaction Scale. Two independent studies were carried out with this purpose. The first study, of 200 participants, analyzed the reliability and factorial structure of the scale with strict criteria of item selection. In the second study, of 275 participants, a confirmatory factorial analysis was carried out and the convergent and concurrent validity of the short version of the instrument (ESSS-B) was calculated. The results of the studies showed adequate psychometric properties in the scale. The reliability index of the scale was of .91 (Cronbach's alpha). The seven items of the ESSS-B complied with the discrimination criteria. Convergent and concurrent validity analyses were performed, obtaining statistically significant values. These results suggest that the ESSS-B has the potential to adequately measure sexual satisfaction in a global and subjective way among Puerto Rican adults.

Keywords: *sexual satisfaction, sexuality, psychometric properties, validation*

*Correspondencia a: Dr. Juan A. González-Rivera, 500 West Main Suite 215, Bayamón, Puerto Rico, 00961. Tel.: 011 787 315 6034. jagonzalez@psm.edu

Nota de Autor: El autor expresa que no hubo conflictos de intereses al redactar el manuscrito.

Cómo citar este artículo: González-Rivera, J. A., & Hernández-Gato, I. (2019). Validación de una versión breve de la Escala de Satisfacción Sexual Subjetiva (ESSS-B) en Puerto Rico. *Revista Evaluar*, 19(2), 43-57. Recuperado de <https://revistas.unc.edu.ar/index.php/revaluar>

Introducción

La salud sexual tiene un rol fundamental en la salud general, en el bienestar y en la calidad de vida de un individuo. La misma ha sido definida como “la integración de los aspectos somáticos, emocionales, intelectuales y sociales de un ser humano sexual en formas que enriquecen y mejoran la personalidad, la comunicación y el amor” (World Health Organization, 2010). A su vez, la satisfacción sexual ha sido identificada como uno de los factores principales de la salud sexual. Asimismo, las dificultades sexuales pueden ocasionar falta de comunicación y conflicto en las parejas, lo cual está altamente relacionado con la disminución en la satisfacción con la relación (Metz & Epstein, 2002; Rowland, van Diest, Incrocci, & Slob, 2005).

Aunque en otras culturas existen instrumentos que permiten evaluar el constructo satisfacción sexual, González-Rivera, Veray-Alicea, Santiago-Santos, Castro-Castro, y Quiñones-Soto (2017) contribuyeron a su investigación científica en Puerto Rico al crear el primer instrumento que permite medir la satisfacción sexual tomando en consideración a la población masculina. En ese momento solo existía el Inventario de Satisfacción Sexual para Mujeres creado por Vizcarrondo-Godreau (2001), el cual limita la evaluación de este constructo a la población femenina puertorriqueña. Los investigadores desarrollaron y validaron la Escala de Satisfacción Sexual Subjetiva (ESSS) con el propósito de ofrecer un indicador general de satisfacción sexual para ser utilizado con fines investigativos y académicos, así como medida de avalúo general en los procesos de psicoterapia.

No obstante, ante la necesidad de tener disponibles instrumentos breves, de fácil administración y con adecuados índices de validez y consistencia interna –como recomienda DeRogatis

(2008)– que favorezcan la medición de la satisfacción sexual en investigaciones académicas, este estudio tiene el propósito de desarrollar una versión breve de la ESSS, ya que un instrumento con estas cualidades facilitaría la tarea investigativa de temas asociados al conflicto y la satisfacción en las relaciones de pareja.

Satisfacción sexual subjetiva

La mayoría de los estudios de satisfacción sexual fueron realizados durante la década de los 90 (Darling, Davidson, & Jennings, 1991; Lawrance & Byers, 1995; Orbuch & Harvey, 1991; Peplau, Cochran, & Mays, 1997; Renaud, Byers, & Pan, 1997; Young, Denny, Luquis, & Young, 1998). A pesar de que el interés en este campo fue aumentando durante las pasadas dos décadas y estudios más recientes han destacado la importancia de la satisfacción sexual en la calidad de vida de los individuos (Ahumada, Lüttges, Molina, & Torres, 2014; Davison, Bell, LaChina, Holden, & Davis, 2009), aún queda mucho por explorar acerca de las implicaciones de este constructo desde una perspectiva holística.

Algunas definiciones de satisfacción sexual tienen una denotación poco precisa. Esto se debe a que los investigadores han utilizado diferentes enfoques para evaluar este constructo y las estrategias de medición varían de acuerdo con los posibles indicadores que podrían afectar la vida sexual de un individuo (Monteiro-Pascoal, Narciso, & Monteiro-Pereira, 2014). Varios modelos teóricos han revelado hallazgos que han sido inconsistentes con los significados operacionales (DeLamater & Hyde, 2004; Sprecher & Cate, 2004). De acuerdo con la estructura de cuatro factores examinada por Hassebrauck y Fehr (2002), la sexualidad representa una de las cuatro dimensiones de la satisfacción de la relación. Los tres facto-

res restantes examinados en este estudio fueron la intimidad, la capacidad para llegar a acuerdos y la independencia. Estos postulados son consistentes con el análisis realizado por Hurlbert, Apt y Rabehl (1993), quienes enfatizaron la importancia de la intimidad y la comunicación en la relación al evaluar la satisfacción sexual en general.

Así como ocurre con muchas definiciones de satisfacción sexual y bienestar, el constructo *satisfacción sexual* también es medido de una manera subjetiva, ya que está respaldado por un conjunto de experiencias personales (McClelland, 2010). Dicho esto, la satisfacción sexual ha sido definida como “una respuesta afectiva que surge de la propia evaluación subjetiva de las dimensiones positivas y negativas asociadas a la propia relación sexual” (Lawrance & Byers, 1995, p. 268). Las personas tienden a calificar sus niveles de satisfacción sexual al analizar el modo en que la actividad sexual cumple con sus expectativas individuales (DeLamater, 1991). Las circunstancias sociales como el sexismo, el estigma sexual y el ambiente cultural son tomadas en consideración al comparar las recompensas sexuales y la percepción de cómo debe ser una relación sexual (Lawrance & Byers, 1995; McClelland, 2010).

La satisfacción sexual ha sido fuertemente vinculada a la satisfacción de la relación, dado que está relacionada con la felicidad que una persona expresa sentir en el aspecto sexual de su relación (Pascoal et al., 2014; Sprecher & Cate, 2004). Algunos estudios han demostrado que las variables *satisfacción de la relación*, *amor* y *compromiso* correlacionan positivamente con la satisfacción sexual tanto en hombres como en mujeres (Sprecher, 2002; Stephenson, Ahrold, & Meston, 2011). Estudios realizados con parejas en relaciones a largo plazo (el 50% de ellas tienen más de 25 años de duración) revelaron discrepancias entre lo manifestado por cada uno de los miembros de la pareja. Las mujeres expresaron sentir-

se más satisfechas sexualmente, mientras que los hombres manifestaron mayor felicidad con la relación (Heiman et al., 2011). A pesar de que los hallazgos relacionados al género no son consistentes, ya que algunos estudios han demostrado diferencias entre hombres y mujeres mientras que otros no, la tendencia a percibir una mayor satisfacción sexual está vinculada con los otros aspectos emocionales que se obtienen al estar comprometidos en una relación (Sánchez-Fuentes, Santos-Iglesias, & Sierra, 2014). Cabe mencionar que se han encontrado diferencias entre los niveles de satisfacción sexual de los individuos que practican el sexo casual y los de quienes que están comprometidos en una relación. Tanto hombres como mujeres han informado niveles más altos de satisfacción sexual al tener una relación estable (Mark, Garcia, & Fisher, 2015).

Por otra parte, algunos estudios de satisfacción sexual han reportado discrepancias al explorar las variables sociodemográficas edad, orientación sexual y nivel de educación. La edad avanzada está relacionada con una menor actividad sexual y, por consiguiente, con una disminución en la satisfacción sexual (De Ryck, Van Laeken, Nöstlinger, Platteau, & Colebunders, 2012; Lindau & Gavrilo, 2010). Otras variables asociadas a la edad son la presencia de enfermedades crónicas y la aparición de disfunciones sexuales (Sierra, Vallejo-Medina, Santos-Iglesias, & Lameiras-Fernández, 2012; Trompeter, Bettencourt, & Barrett-Connor, 2012). En términos de la orientación sexual, algunos estudios asocian la homosexualidad con un aumento en la satisfacción sexual (Henderson, Lehavot, & Simoni, 2009), mientras que los hallazgos de otras investigaciones han revelado una mayor satisfacción sexual en hombres heterosexuales (Dixon, 1985). Niveles más altos de satisfacción sexual también han sido correlacionados con un alto nivel de educación (Carpenter, Nathanson, & Kim, 2009) y

menor cantidad de parejas sexuales (Heiman et al., 2011).

Asimismo, se han realizado investigaciones para asociar la importancia de la imagen corporal con el funcionamiento sexual. Estos estudios han sido mayormente enfocados en mujeres dado el énfasis otorgado a la imagen de la mujer, particularmente en las culturas occidentales. Los hallazgos de varios estudios han confirmado que las mujeres que se sienten más cómodas con su imagen corporal revelan niveles más altos de experiencia sexual, mayor frecuencia en la actividad sexual y mayor disposición para iniciar los encuentros sexuales, comparado a las que se sienten menos satisfechas con su imagen (Ackard, Kearney-Cooke, & Peterson, 2000; Schooler, Ward, Merriwether, & Caruthers, 2005). Una percepción negativa de la imagen corporal en la mujer también ha sido relacionada con la evitación del encuentro sexual. El estudio de La Rocque y Cioe (2011) evaluó esta asociación y confirmó que las mujeres con una autoimagen negativa presentaron una mayor evasión ante la actividad sexual. Del mismo modo, el que las mujeres no se sientan a gusto con su imagen contribuye a la disminución del placer, los orgasmos y la satisfacción sexual (Sanchez & Kiefer, 2007; Yamamiya, Cash, & Thompson, 2006).

Cabe mencionar que el orgasmo y su consistencia han sido utilizados como medidas en los estudios de satisfacción, ya que ambas han ofrecido datos cuantificables y válidos a través de los participantes (Darling et al., 1991; Haavio-Mannila & Kontula, 1997; Young, Denny, Young, & Luquis, 2000). Mientras que el orgasmo puede proveer un cierto tipo de resultado acerca de la experiencia sexual, es solo una faceta de esta y no necesariamente está asociado a la satisfacción en todos los individuos (Haavio-Mannila & Kontula, 1997). De acuerdo con Hurlbert et al. (1993), el deseo, la excitación y la consistencia del orgasmo

están vinculados a una mayor satisfacción sexual. No obstante, las mujeres pueden expresar niveles más bajos de frecuencia orgásmica e incluso sentirse satisfechas sexualmente (McClelland, 2009). Algunos individuos también han revelado mayores niveles de orgasmo durante el coito cuando tienen un encuentro sexual con sus parejas, comparado con los que tienen encuentros sexuales fuera de relaciones comprometidas (Armstrong, England, & Fogarty, 2012; Mark et al., 2015). La frecuencia del coito y la libido también han correlacionado positivamente con el constructo satisfacción sexual (Daker-White & Donovan, 2002; Young et al., 1998).

Escala de Satisfacción Sexual Subjetiva (ESSS) – Versión original

González-Rivera et al. (2017) desarrollaron y validaron la versión original de la ESSS. Esta escala fue desarrollada por sus investigadores con el propósito de medir el constructo satisfacción sexual desde un modelo de cuatro factores: *valoración subjetiva, aspecto emocional, ejecución sexual y autoimagen*. La escala fue administrada en una muestra de 615 adultos puertorriqueños de entre 20 y 75 años. Las puntuaciones de la ESSS tienen un rango de 20 a 80, siendo las mayores puntuaciones las que revelan una mayor satisfacción sexual en la persona. La versión original de la ESSS consta de 20 reactivos (5 en cada dimensión), los cuales reflejan indicadores adecuados de validez y confiabilidad. Para examinar la validez del instrumento se realizaron análisis correlacionales entre la ESSS y los cuatro factores y se obtuvieron resultados aceptables. Los análisis de consistencia interna para determinar la confiabilidad de la escala original mostraron índices adecuados de .91 y .89, según las pruebas alfa de Cronbach y Spearman-Brown.

El objetivo principal de este proyecto es desarrollar una versión breve de la ESSS que facilite la tarea investigativa de temas asociados al conflicto y la satisfacción en las relaciones de pareja. Para ello, se llevaron a cabo dos estudios independientes. En el Estudio 1, con la intención de desarrollar una versión breve de la ESSS, se analiza la estructura factorial de la escala mediante un análisis factorial exploratorio (AFE) con criterios estrictos de selección, se evalúa la confiabilidad del instrumento y se examina la capacidad de discriminación de los ítems. En el Estudio 2, se analiza la validez de constructo de la versión breve de la escala mediante un análisis factorial confirmatorio (AFC) con ecuaciones estructurales, se evalúa la confiabilidad compuesta y se examina la validez concurrente y convergente del instrumento. En los dos estudios se utiliza un diseño de investigación instrumental (Montero & León, 2007). Ambos estudios fueron aprobados por el Comité para la Ética en la Investigación (IRB) de la Universidad Carlos Albizu en San Juan, Puerto Rico.

Estudio 1

Método

Participantes

Para participar del Estudio 1, establecimos los siguientes criterios de inclusión para adecuar la muestra al propósito del estudio (validar un instrumento que facilite estudiar temas asociados al conflicto y la satisfacción en las relaciones de pareja): tener 21 años o más y estar en una relación de pareja (matrimonio, convivencia libre o noviazgo). Se trabajó con una muestra no probabilística constituida por 200 participantes, de los cuales 157 eran mujeres (78.5%) y 43 hombres (21.5%). El rango de edad estaba entre los 21 y los 60 años, de los cuales el 31.5% (n = 63) tenía

entre 21 y 30 años, el 30.5% (n = 61) entre 31 y 40 años, el 34% (n = 68) entre 41 y 50 años y el 4% (n = 8) tenía más de 51 años. En cuanto al estado civil, el 69.5% (n = 174) estaba casado legalmente, el 24% (n = 48) convivía con su pareja sin nupcias y el 6.5% (n = 13) era soltero, pero en una relación de noviazgo. En términos de ingresos económicos, el 44.5% (n = 89) generaba ingresos menores a \$25'000, mientras que el 32.5% (n = 65) entre \$26'000 y \$50'000, y el 23% (n = 46) ingresos superiores a \$50'000. Por último, el 98% (n = 196) se identificó como heterosexual, 1% (n = 2) como bisexual y 1% (n = 2) como lesbiana.

Instrumento

La Escala de Satisfacción Sexual Subjetiva (ESSS), desarrollada por González-Rivera et al. (2017), se compone de 20 ítems organizados en una escala tipo Likert de cuatro puntos que va de (1) *Totalmente en desacuerdo* a (4) *Totalmente de acuerdo*. La escala consta de cuatro factores: valoración subjetiva, aspecto emocional, ejecución sexual y autoimagen. El rango posible es de 20 a 80 puntos. A mayor puntuación, mayor satisfacción sexual exhibe la persona ($\alpha = .90$).

Procedimientos generales

La recopilación de datos se llevó a cabo mediante el uso de cuestionarios a través de la plataforma PsychData utilizando como método de reclutamiento las redes sociales Facebook, Twitter, Google+ y WhatsApp, entre otras. Esto permitió un efecto bola de nieve en las redes sociales. Una vez que accedían a la encuesta en línea, los participantes leían el consentimiento informado, en el cual se notificaba el propósito del estudio, los criterios de inclusión, su naturaleza voluntaria

y los posibles riesgos y beneficios, así como su derecho a retirarse del estudio en cualquier momento. Además, se les notificaba sobre la duración de su participación y su derecho a obtener los resultados del estudio una vez culminara la investigación. Para garantizar la privacidad y confidencialidad de los participantes, los cuestionarios se completaban de forma anónima y los participantes tenían la posibilidad de imprimir una copia del consentimiento informado. Los datos fueron analizados en el programa estadístico IBM SPSS v. 24 (IBM Corp., 2016).

Resultados

Análisis factorial

En el primer estudio se realizó un AFE utilizando el método de extracción de máxima verosimilitud con rotación oblicua. Como criterios de aceptación, consideramos aquellos ítems con una carga factorial mayor a .50 en un solo factor (Stevens, 2002). Todos los ítems que no cumplieran este criterio serían eliminados de la versión corta del instrumento. Los resultados mostraron una solución factorial que explica el 64% de la varianza de los datos ($KMO = .89$; $\chi^2 = 1787.963_{(190)}$, $p < .001$). En este primer análisis, de los 20 ítems de la escala original, se eliminaron siete ítems por no cumplir con los criterios de selección (ítems 6, 10, 11, 12, 16, 19 y 20). La versión de 13 ítems fue sometida a un segundo AFE que mostró una solución factorial que explica el 65% de la varianza de los datos ($KMO = .90$; $\chi^2 = 1253.05_{(78)}$, $p < .001$). En esta ocasión, dos ítems no cumplieron con los criterios de selección, por lo cual fueron eliminados (ítems 4 y 15).

Un tercer AFE con la versión de 11 ítems mostró una solución factorial que explica el 62% de la varianza de los datos ($KMO = .92$; $\chi^2 = 1082.998_{(55)}$, $p < .001$). En esta ocasión, otros

cuatro ítems no cumplieron con los criterios de selección (ítems 7, 8, 13 y 18). La última versión de siete ítems fue sometida a un AFE. Los resultados confirmaron que estos siete ítems cargaron en un solo factor, cumplieron con los criterios de selección y explicaron el 65% de la varianza de los datos ($KMO = .92$; $\chi^2 = 792.128_{(21)}$, $p < .001$). En la Tabla 1 se presentan las cargas factoriales obtenidas por los siete ítems de la versión final.

Análisis de Discriminación

Una vez que se obtuvo una versión reducida de la escala, se llevó a cabo un análisis de reactivos para calcular los índices de discriminación de los siete ítems de la versión breve mediante el método de correlación del ítem con el total de la prueba (r_{bis}). Los resultados mostraron índices de discriminación que fluctuaban de .67 a .81, de manera que todos los ítems obtuvieron índices de discriminación mayores a .30, como recomienda Kline (2005). La Tabla 1 presenta los índices de discriminación de los siete ítems de la escala.

Análisis de Confiabilidad

La versión breve de la escala fue sometida a un análisis de consistencia interna para determinar el índice de confiabilidad de la escala. Para ello utilizamos dos métodos: el coeficiente alfa de Cronbach y división en mitades de Spearman-Brown. Los resultados mostraron un coeficiente alfa de .91 y de .87 para la prueba Spearman-Brown. La Tabla 2 presenta el alfa de Cronbach, alfa de Cronbach estandarizado, el coeficiente de la prueba Spearman-Brown, la media y la desviación estándar de la Versión Breve de la Escala de Satisfacción Sexual Subjetiva (ESSS-B).

Tabla 1

Análisis de factores exploratorios e índices de discriminación de los ítems.

Ítems	r_{bis}	1
1. Considero que mi vida sexual es muy excitante.	.77	.83
2. Luego de tener relaciones sexuales me siento pleno.	.69	.73
3. Estoy satisfecho con la cantidad de sexo que practico a la semana.	.73	.79
5. Mi vida sexual es divertida.	.81	.86
9. A mi vida sexual le falta calidad.	.67	.70
14. Me siento satisfecho con la frecuencia de mis orgasmos.	.67	.70
17. Me siento cómodo con la calidad de sexo que practico.	.71	.74

Nota. r_{bis} = índice de discriminación del ítem; 1 = cargas factoriales. (n = 200).

Tabla 2

Media, desviación estándar y coeficientes de confiabilidad.

Factor	M	DE	α	α_{est}	Spearman-Brown
ESSS- B (siete ítems)	20.56	5.33	.91	.91	.87

Nota. M = media; DE = desviación estándar; α = alfa de Cronbach; α_{est} = alfa de Cronbach estandarizado. (n = 200).

Estudio 2

Método

Participantes

Para participar del Estudio 2, establecimos los mismos criterios de inclusión que para el estudio 1: tener 21 años o más y estar en una relación de pareja, dado que es requerimiento de uno de los instrumentos administrados (Escala de Evaluación de la Relación de Pareja). Se trabajó con una muestra no probabilística constituida por 275 participantes. De estos, 212 eran mujeres (77.1%) y 63 hombres (22.9%). El rango de edad estaba entre los 21 y los 62 años, con una edad promedio de 32.80 (DE = 6.99). En cuanto al estado civil, el 48.4% (n = 133) estaba casado legalmente y el 51.6% (n = 142) convivía con su pareja sin nupcias. En términos de la preparación académica, el 38.9% (n = 107) poseía un bachillerato,

el 21.5% (n = 59) poseía un grado de maestría, el 25.8% (n = 71) tenía un grado técnico o grado asociado, el 8.4% (n = 23) completó únicamente la escuela superior y el 5.5% (n = 15) contaba con un grado doctoral. La mayoría de la muestra (64.7%, n = 178) generaba ingresos anuales menores a \$25'000, el 27.3% (n = 75) entre \$26'000 y \$50'000 y el 8% (n = 22) ingresos superiores a \$50'000 anuales. Por último, el 100% (n = 275) se identificó como heterosexual.

Instrumentos

En el Estudio 2 se utilizaron tres instrumentos: la Escala de Satisfacción Sexual Subjetiva (ESSS; véase descripción en el Estudio 1), la versión breve de la Escala de Satisfacción Sexual Subjetiva (ESSS-B) y la Escala de Evaluación de

la Relación de Pareja. La ESSS-B, validada en el Estudio 1, se compone de 7 ítems organizados en una escala tipo Likert de cuatro puntos que va de (1) *Totalmente en desacuerdo* a (4) *Totalmente de acuerdo*. El rango posible es de 7 a 28 puntos. A mayor puntuación, mayor satisfacción sexual reporta la persona ($\alpha = .91$).

Por su parte, la Escala de Evaluación de la Relación de Pareja (Hendrick, 1988) mide de forma global los sentimientos, pensamientos o conductas en la relación de pareja (por ejemplo, *¿Qué tanto tu pareja satisface tus necesidades?*). Consta de siete ítems con una escala de respuesta de cinco puntos que fluctúan de *Nada a Totalmente*. El rango posible es de 7 a 35 puntos. A mayor puntaje, mayor satisfacción y calidad en la relación de pareja exhibe la persona ($\alpha = .92$).

Procedimientos generales

La recopilación de datos se llevó a cabo mediante el uso de cuestionarios a través de la plataforma PsychData utilizando como método de reclutamiento las redes sociales Facebook, Twitter, Google+ y WhatsApp, entre otras. Esto permitió un efecto bola de nieve en las redes sociales. Una vez que accedían a la encuesta en línea, los participantes leían el consentimiento informado, en el cual se notificaba el propósito del estudio, los criterios de inclusión, su naturaleza voluntaria y los posibles riesgos y beneficios, así como su derecho a retirarse del estudio en cualquier momento. Además, se les notificaba sobre la duración de su participación y su derecho a obtener los resultados del estudio una vez que culminara la investigación. Para garantizar la privacidad y confidencialidad de los participantes, los cuestionarios se completaron de forma anónima y los participantes tenían la posibilidad de imprimir una copia del consentimiento informado. Los datos fueron

analizados en el programa estadístico STATA 15 (StataCorp., 2017).

Resultados

Análisis descriptivos de los ítems

En primer lugar, se calcularon las medias y desviaciones estándar para cada ítem de la versión breve para analizar la distribución de la escala. Las medias de los ítems fluctuaron entre 2.44 y 2.99, con desviaciones estándar que fluctuaron entre .90 y 1.13. Como muestra la Tabla 3, las pruebas de normalidad Kolmogorov-Smirnov y Shapiro-Wilk evidencian que los datos no se distribuyen de forma normal ($p < .001$). Por la falta de normalidad en los datos, utilizamos las correcciones de Satorra y Bentler (2001) para calcular el ajuste de los modelos de ecuaciones estructurales, ya que la no normalidad de los datos altera los errores de estimación y el ajuste global del modelo.

Validez de Constructo

Se realizó un AFC con ecuaciones estructurales utilizando el método de estimación de máxima verosimilitud. Para evaluar el ajuste del modelo se utilizaron los siguientes índices de bondad de ajuste y los criterios de Byrne (2010): Chi-cuadrado (χ^2), error medio cuadrático de aproximación (RMSEA; valores iguales o menores a .08 indican un buen ajuste), índice Tucker-Lewis (TLI) y el índice de ajuste comparativo (CFI). Para que exista un buen ajuste del modelo, los valores sugeridos de CFI y TLI deben ser mayores o iguales que .95 (Byrne, 2010). Los resultados arrojaron un buen ajuste para el modelo conformado por un solo factor latente que recogía los siete ítems de la escala (véase Figura 1),

Tabla 3

Estadísticos de descriptivos y de distribución para los ítems de la escala.

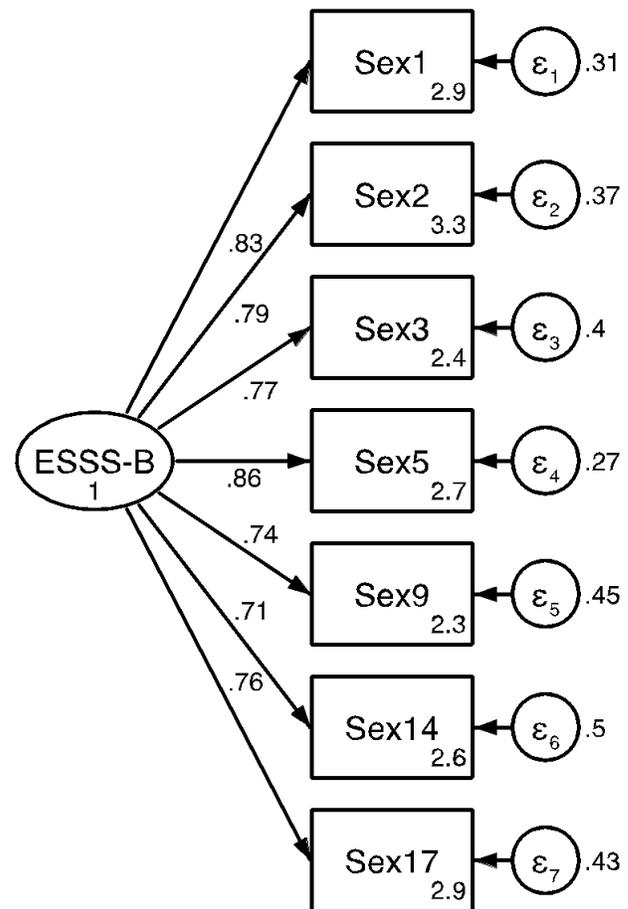
Ítem	Media	Desviación Estándar	Asimetría	Curtosis	Kolmogorov-Smirnov	Shapiro-Wilk	r_{bis}
1	2.68	0.94	-0.30	-0.77	.25	.870	.78
2	2.99	0.90	-0.55	-0.52	.23	.845	.75
3	2.44	1.00	0.09	-1.06	.21	.877	.73
5	2.61	0.97	-0.06	-0.97	.20	.878	.80
9	2.55	1.13	0.02	-1.39	.21	.847	.71
14	2.66	1.01	-0.26	-1.02	.23	.869	.69
17	2.75	0.96	-0.26	-0.90	.22	.871	.73

Nota. r_{bis} = índice de discriminación del ítem; error estándar de la asimetría = .15; error estándar de la curtosis = .29; grados de libertad Kolmogorov-Smirnov y Shapiro-Wilk = 275; todos los valores $p < .001$.

$\chi^2 = 44.845_{(14)}$, $p < .001$; RMSEA = .09; CFI = .97, TLI = .96; χ^2 corregido = $36.693_{(14)}$, $p < .001$; RMSEA corregido = .08; CFI corregido = .98, TLI corregido = .97. A su vez, se examinaron los coeficientes de regresión de cada ítem; se esperaba que cada uno de estos fuese mayor a .65, según recomiendan Schumaker y Lomax (2010). Los coeficientes de regresión fluctuaron entre .71 y .86 (véase Tabla 4). Luego, se analizó la confiabilidad compuesta del modelo unidimensional de la escala. Los análisis revelaron un índice de confiabilidad compuesta de .92 y una varianza media explicada de .61, superando el valor mínimo recomendado por la literatura (Raykov & Shrout, 2002).

Validez de Criterio

Con el fin de evaluar la validez concurrente y convergente de la ESSS-B, se analizaron las correlaciones con la versión larga del instrumento y una medida de satisfacción en la relación de pareja. Como se puede observar en la Tabla 5, todas las correlaciones fueron significativas. Según lo esperado, las correlaciones entre las dos versio-

**Figura 1**

Modelo unidimensional de la Versión Breve de la Escala de Satisfacción Sexual Subjetiva (ESSS-B).

Tabla 4

Estimados estandarizados de regresión y sus intervalos de confianza.

Ítems	β	p	I.C. 95%
1. Considero que mi vida sexual es muy excitante.	.83	< .001	[.79, .88]
2. Luego de tener relaciones sexuales me siento pleno.	.79	< .001	[.63, .76]
3. Estoy satisfecho con la cantidad de sexo que practico a la semana.	.77	< .001	[.75, .84]
5. Mi vida sexual es divertida.	.86	< .001	[.80, .92]
9. A mi vida sexual le falta calidad.	.74	< .001	[.66, .82]
14. Me siento satisfecho con la frecuencia de mis orgasmos.	.71	< .001	[.64, .78]
17. Me siento cómodo con la calidad de sexo que practico.	.76	< .001	[.70, .82]

Nota. β = estimados estandarizados de regresión; p = significación; I.C. 95% = intervalos de confianza de los coeficientes de regresión.

Tabla 5

Correlaciones entre las escalas, sus medias, desviaciones estándar y alfa de Cronbach.

	α	M	DE	1	2
1. ESSS-B	.91	18.68	5.64		
2. ESSS (versión larga)	.90	58.72	10.81	.89	
3. Satisfacción en la Relación	.90	25.59	5.73	.68	.61

Nota. Todas las correlaciones significativas a .001. (n = 342); α = alfa de Cronbach.

nes del instrumento fueron significativamente altas y positivas. Asimismo, la correlación entre la ESSS-B y la satisfacción en la relación de pareja fue positiva.

Discusión

El presente proyecto de investigación tuvo como propósito desarrollar una versión breve de la ESSS que facilite la tarea investigativa de temas asociados al conflicto y a la satisfacción en las relaciones de pareja. Para ello, se analizaron las propiedades psicométricas de la ESSS utilizando criterios estrictos de selección de ítems en dos muestras de adultos puertorriqueños. A partir de los resultados obtenidos en ambos estudios, se obtuvo una versión breve del instrumento con

siete ítems (ESSS-B) que posee adecuadas características psicométricas para examinar de manera subjetiva y general la satisfacción sexual de una persona. Asimismo, los índices de confiabilidad obtenidos sugieren, según lo establece [DeVellis \(2017\)](#), que la ESSS-B cuenta con la suficiente consistencia interna para ser utilizada como instrumento de medición científica en investigaciones en Puerto Rico e incluso como instrumento de avalúo general en los procesos de psicoterapia y consejería.

De forma general, el adiconamiento del AFE del Estudio 1 con el AFC del Estudio 2 demuestra que la versión unidimensional de la ESSS-B presenta un ajuste satisfactorio con los datos. Esto es congruente con las tendencias actuales del estudio de la satisfacción sexual que favorecen la medición de este constructo desde una

manera subjetiva basada en el juicio global de las experiencias vividas por las personas (Lawrance & Byers, 1995; McClelland, 2010). En otras palabras, las puntuaciones obtenidas en la ESSS-B cubren varios aspectos de la experiencia sexual de una forma global, sobre los cuales la persona emite un juicio valorativo, ya sea positivo (satisfacción) o negativo (insatisfacción). Asimismo, pueden considerarse como una actitud de favorabilidad o desfavorabilidad hacia su actividad sexual al contrastarlas con sus expectativas individuales.

En cuanto a la confiabilidad del instrumento, se obtuvieron índices superiores a los recomendados por la literatura (Campo-Arias & Oviedo, 2008; Kline, 2005), lo cual sugiere que la versión final de la escala es un instrumento estable, reproducible y consistente en la medida de la satisfacción sexual. De igual forma, las correlaciones de cada ítem con la puntuación total manifiestan una consistencia interna notable. Esto sugiere que los ítems de la versión final discriminan adecuadamente y son capaces de diferenciar personas con distintos niveles de satisfacción sexual.

Por su parte, la correlación moderadamente alta de la ESSS-B con la medida de satisfacción en la pareja aporta evidencia sobre la validez convergente de la escala y concuerda con los patrones de correlaciones obtenidos en otras investigaciones (Monteiro-Pascoal et al., 2014; Sprecher & Cate, 2004; Stephenson et al., 2011). De igual forma, la correlación alta de las dos versiones del instrumento confirma que la versión ESSS-B tiene la capacidad de cuantificar de forma significativa y adecuada el mismo constructo que evalúa la ESSS. Esto demuestra que la ESSS-B es una medida confiable de satisfacción sexual y aporta evidencia a su validez concurrente.

En cuanto a las implicaciones prácticas del estudio, se demostró que la ESSS-B es útil para el desarrollo de nuevas investigaciones en el campo

de la psicología y la sexualidad en Puerto Rico. En términos de conveniencia, la principal fortaleza de la escala es su lenguaje sencillo y su brevedad, lo que la consolida como un instrumento de fácil administración e interpretación. En este sentido, la ESSS-B se constituye como una herramienta práctica y efectiva en el quehacer investigativo y académico de la psicología. Las puntuaciones de la ESSS-B se calculan mediante la sumatoria de los siete ítems para obtener un índice general. Puntuaciones entre 25 y 28 se consideran indicadores de una alta satisfacción sexual; puntuaciones entre 15 y 24 se consideran indicadores de una satisfacción sexual promedio; y puntuaciones entre 7 y 14 indican una baja satisfacción sexual.

Al igual que toda investigación, nuestro estudio no está exento de limitaciones. Primero, la muestra fue reunida de forma accidental y no fue aleatoria (muestra no probabilística). Segundo, no se pudo establecer la confiabilidad del instrumento a través del tiempo, solo se pudo hacer a través de sus elementos. No obstante, las técnicas avanzadas utilizadas en el estudio brindan robustez empírica a nuestros resultados. Tercero, las muestras utilizadas en ambos estudios están integradas por personas con parejas (casados, unión libre, noviazgo), por lo cual no sabemos cómo se comporta la escala en personas sin pareja que mantienen una vida sexual activa. Cuarto, el 98% de los participantes del Estudio 1 y la totalidad de la muestra del Estudio 2 se identificaron como heterosexuales, por lo cual nuestros hallazgos y conclusiones se limitan a personas heterosexuales con pareja. Por último, el procedimiento para recoger los datos no fue estandarizado, lo que puede afectar las medias del estudio y aumentar el error estándar de medición.

Para futuros estudios, se recomienda examinar la confiabilidad temporal a través de la técnica de test-retest y realizar el proceso de validación

cruzada. A su vez, se recomienda que se valide la ESSS-B en otras poblaciones latinoamericanas para examinar sus propiedades psicométricas en distintos contextos nacionales e internacionales. Por último, se sugiere replicar estos resultados en muestras más amplias, equilibradas por sexo y que incluyan un número considerable de participantes de la comunidad LGBT.

En síntesis, las propiedades psicométricas obtenidas demuestran que la ESSS-B refleja indicadores adecuados de validez y confiabilidad. El hecho de utilizar dos muestras independientes para validar y examinar la ESSS-B da un valor añadido al estudio. Se concluye que la ESSS-B es un excelente instrumento de medición para futuras investigaciones en Puerto Rico y recomendamos su validación en otros países latinoamericanos.

Referencias

- Ackard, D. M., Kearney-Cooke, A., & Peterson, C. B. (2000). Effects of body image and self-image on women's sexual behaviors. *International Journal of Eating Disorders, 28*(4), 422-429. doi: [10.1002/1098-108x\(200012\)28:4<422::aid-eat10>3.0.co;2-1](https://doi.org/10.1002/1098-108x(200012)28:4<422::aid-eat10>3.0.co;2-1)
- Ahumada, S., Lüttges, C., Molina, T., & Torres, S. (2014). Satisfacción sexual: Revisión de los factores individuales y de pareja relacionados. *Revista Hospital Clínico Universidad de Chile, 25*(4), 278-284. Recuperado de <https://www.redclinica.cl>
- Armstrong, E. A., England, P., & Fogarty, A. C. K. (2012). Accounting for women's orgasm and sexual enjoyment in college hookups and relationships. *American Sociological Review, 77*(3), 435-462. doi: [10.1177/0003122412445802](https://doi.org/10.1177/0003122412445802)
- Byrne, B. M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (2ª ed.). New York, NY: Taylor & Francis Group. doi: <https://www.taylorfrancis.com/books/9781410600219>
- Campo-Arias, A., & Oviedo, H. C. (2008). Propiedades psicométricas de una escala: La consistencia interna. *Revista de Salud Pública, 10*(5), 831-839. doi: [10.1590/s0124-00642008000500015](https://doi.org/10.1590/s0124-00642008000500015)
- Carpenter, L. M., Nathanson, C. A., & Kim, Y. J. (2009). Physical women, emotional men: Gender and sexual satisfaction in midlife. *Archives of Sexual Behavior, 38*(1), 87-107. doi: [10.1007/s10508-007-9215-y](https://doi.org/10.1007/s10508-007-9215-y)
- Daker-White, G., & Donovan, J. (2002). Sexual satisfaction, quality of life and the transaction of intimacy in hospital patients' accounts of their (hetero)sexual relationships. *Sociology of Health and Illness, 24*(1), 89-113. doi: [10.1111/1467-9566.00005](https://doi.org/10.1111/1467-9566.00005)
- Darling, C. A., Davidson, J. K., & Jennings, D. A. (1991). The female sexual response revisited: Understanding the multiorgasmic experience in women. *Archives of Sexual Behavior, 20*(6), 527-540. doi: [10.1007/BF01550952](https://doi.org/10.1007/BF01550952)
- Davison, S. L., Bell, R. J., LaChina, M., Holden, S. L., & Davis, S. R. (2009). The relationship between self-reported sexual satisfaction and general well-being in women. *The Journal of Sexual Medicine, 6*(10), 2690-2697. doi: [10.1111/j.1743-6109.2009.01406.x](https://doi.org/10.1111/j.1743-6109.2009.01406.x)
- De Ryck, I., van Laeken, D., Nöstlinger, C., Platteau, T., & Colebunders, R. (2012). Sexual satisfaction among men living with HIV in Europe. *AIDS and Behavior, 16*(1), 225-230. doi: [10.1007/s10461-011-9987-x](https://doi.org/10.1007/s10461-011-9987-x)
- DeLamater, J. (1991). Emotions and sexuality. En K. McKinney & S. Sprecher (Eds.), *Sexuality in Close Relationships* (pp. 49-70). Hillsdale, NJ: Lawrence Erlbaum.
- DeLamater, J., & Hyde, J. (2004). Conceptual and theoretical issues in studying sexuality in close relationships. En J. H. Harvey, A. Wenzel, & S. Sprecher (Eds.), *The Handbook of Sexuality in Close Relationships* (pp. 7-30). Mahwah, NJ: Lawrence Erlbaum. doi: [10.4324/9781410610249](https://doi.org/10.4324/9781410610249)
- DeRogatis, L. R. (2008). Assessment of sexual function/dysfunction via patient reported outcomes. *International Journal of Impotence Research, 20*(1),

- 35-44. doi: [10.1038/sj.jjir.3901591](https://doi.org/10.1038/sj.jjir.3901591)
- DeVellis, R. F. (2017). *Scale development: Theory and applications*. California, CA: Sage.
- Dixon, D. (1985). Perceived sexual satisfaction and marital happiness of bisexual and heterosexual swinging husbands. *Journal of Homosexuality*, *11*(1-2), 209-222. doi: [10.1300/J082v11n01_15](https://doi.org/10.1300/J082v11n01_15)
- González-Rivera, J. A., Veray-Alicea, J., Santiago-Santos, D., Castro-Castro, S., & Quiñones-Soto, R. (2017). Desarrollo y validación de una escala para medir satisfacción sexual subjetiva en adultos puertorriqueños. *Salud y Conducta Humana*, *4*(1), 52-63. Recuperado de <http://rsych.com>
- Haavio-Mannila, E., & Kontula, O. (1997). Correlates of increased sexual satisfaction. *Archives of Sexual Behavior*, *26*(4), 399-419. Recuperado de <https://link.springer.com/journal/volumesAndIssues/10508>
- Hassebrauck, M., & Fehr, B. (2002). Dimensions of relationship quality. *Personal Relationships*, *9*(3), 253-270. doi: [10.1111/1475-6811.00017](https://doi.org/10.1111/1475-6811.00017)
- Heiman, J. R., Long, J. S., Smith, S. N., Fisher, W. A., Sand, M. S., & Rosen, R. C. (2011). Sexual satisfaction and relationship happiness in midlife and older couples in five countries. *Archives of Sexual Behavior*, *40*(4), 741-753. doi: [10.1007/s10508-010-9703-3](https://doi.org/10.1007/s10508-010-9703-3)
- Henderson, A. W., Lehavot, K., & Simoni, J. M. (2009). Ecological models of sexual satisfaction among lesbian/bisexual and heterosexual women. *Archives of Sexual Behavior*, *38*(1), 50-65. doi: [10.1007/s10508-008-9384-3](https://doi.org/10.1007/s10508-008-9384-3)
- Hendrick, S. S. (1988). A generic measure of relationship satisfaction. *Journal of Marriage and Family*, *50*(1), 93-98. doi: [10.2307/352430](https://doi.org/10.2307/352430)
- Hurlbert, D. F., Apt, C., & Rabehl, S. M. (1993). Key variables to understanding female sexual satisfaction: An examination of women in nondistressed marriages. *Journal of Sex & Marital Therapy*, *19*(2), 154-165. doi: [10.1080/00926239308404899](https://doi.org/10.1080/00926239308404899)
- IBM (2016). IBM SPSS Statistics for Windows, Version 24.0. [software de cómputo]. Armonk, NY: IBM.
- Kline, T. J. B. (2005). *Psychological testing: A practical approach to design and evaluation*. Thousand Oaks, CA: Sage. doi: [10.4135/9781483385693](https://doi.org/10.4135/9781483385693)
- La Rocque, C. L., & Cioe, J. (2011). An evaluation of the relationship between body image and sexual avoidance. *The Journal of Sex Research*, *48*(4), 397-408. doi: [10.1080/00224499.2010.499522](https://doi.org/10.1080/00224499.2010.499522)
- Lawrance, K. A., & Byers, E. S. (1995). Sexual satisfaction in long-term heterosexual relationships: The interpersonal exchange model of sexual satisfaction. *Personal Relationships*, *2*(4), 267-285. doi: [10.1111/j.1475-6811.1995.tb00092.x](https://doi.org/10.1111/j.1475-6811.1995.tb00092.x)
- Lindau, S. T., & Gavrilo, N. (2010). Sex, health, and years of sexually active life gained due to good health: Evidence from two US population based cross sectional surveys of ageing. *BMJ: British Medical Journal*, 1-11. doi: [10.1136/bmj.c810](https://doi.org/10.1136/bmj.c810)
- Mark, K. P., Garcia, J. R., & Fisher, H. E. (2015). Perceived emotional and sexual satisfaction across sexual relationship contexts: Gender and sexual orientation differences and similarities. *The Canadian Journal of Human Sexuality*, *24*(2), 120-130. doi: [10.3138/cjhs.242-A8](https://doi.org/10.3138/cjhs.242-A8)
- McClelland, S. I. (2009). *Intimate justice: Sexual satisfaction in young adults* (Disertación doctoral). Recuperado de Tesis y Disertaciones de ProQuest. (Número UMI 3365820)
- McClelland, S. I. (2010). Intimate justice: A critical analysis of sexual satisfaction. *Social and Personality Psychology Compass*, *4*(9), 663-680. doi: [10.1111/j.1751-9004.2010.00293.x](https://doi.org/10.1111/j.1751-9004.2010.00293.x)
- Metz, M. E., & Epstein, N. (2002). Assessing the role of relationship conflict in sexual dysfunction. *Journal of Sex and Marital Therapy*, *28*(2), 139-164. doi: [10.1080/00926230252851889](https://doi.org/10.1080/00926230252851889)
- Monteiro-Pascoal, P., Narciso, I. de S. B., & Monteiro-Pereira, N. (2014). What is sexual satisfaction? Thematic analysis of lay people's definitions. *The Journal of Sex Research*, *51*(1), 22-30. doi: [10.1080/00224499.2013.815149](https://doi.org/10.1080/00224499.2013.815149)
- Montero, I., & León, O. G. (2007). A guide for naming research studies in psychology. *International Journal*

- of *Clinical and Health Psychology*, 7(3), 847-862.
Recuperado de <http://www.aepc.es>
- Orbuch, T. L., & Harvey, J. H. (1991). Methodological and conceptual issues in the study of sexuality in close relationships. En K. McKinney & S. Sprecher (Eds.), *Sexuality in Close Relationships* (pp. 9-24). Hillsdale, NJ: Erlbaum.
- Peplau, L. A., Cochran, S. D., & Mays, V. M. (1997). A national survey of the intimate relationships of African American lesbians and gay men: A look at commitment, satisfaction, sexual behavior, and HIV disease. En B. Greene (Ed.), *Ethnic and cultural diversity among lesbians and gay men. Serie Psychological perspectives on lesbian and gay issues* (Vol. 3, pp. 11-38). Thousand Oaks, CA: Sage.
- Raykov, T., & Shrout, P. E. (2002). Reliability of scales with general structure: Point and interval estimation using a structural equation modeling approach. *Structural Equation Modeling: A Multidisciplinary Journal*, 9(2), 195-212. doi: [10.1207/S15328007SEM0902_3](https://doi.org/10.1207/S15328007SEM0902_3)
- Renaud, C., Byers, E. S., & Pan, S. (1997). Sexual and relationship satisfaction in mainland China. *The Journal of Sex Research*, 34(4), 399-410. doi: [10.1080/00224499709551907](https://doi.org/10.1080/00224499709551907)
- Rowland, D., Van Diest, S., Incrocci, L., & Slob, A. K. (2005). Psychosexual factors that differentiate men with inhibited ejaculation from men with no dysfunction or another sexual dysfunction. *The Journal of Sexual Medicine*, 2(3), 383-389. doi: [10.1111/j.1743-6109.2005.20352.x](https://doi.org/10.1111/j.1743-6109.2005.20352.x)
- Sánchez-Fuentes, M. del M., Santos-Iglesias, P., & Sierra, J. C. (2014). A systematic review of sexual satisfaction. *International Journal of Clinical and Health Psychology*, 14(1), 67-75. doi: [10.1016/S1697-2600\(14\)70038-9](https://doi.org/10.1016/S1697-2600(14)70038-9)
- Sanchez, D. T., & Kiefer, A. K. (2007). Body concerns in and out of the bedroom: Implications for sexual pleasure and problems. *Archives of Sexual Behavior*, 36(6), 808-820. doi: [10.1007/s10508-007-9205-0](https://doi.org/10.1007/s10508-007-9205-0)
- Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, 66(4), 507-514. doi: [10.1007/BF02296192](https://doi.org/10.1007/BF02296192)
- Schooler, D., Ward, L. M., Merriwether, A., & Caruthers, A. S. (2005). Cycles of shame: Menstrual shame, body shame, and sexual decision-making. *The Journal of Sex Research*, 42(4), 324-334. doi: [10.1080/00224490509552288](https://doi.org/10.1080/00224490509552288)
- Schumacker, R. E., & Lomax, R. G. (2010). *A beginner's guide to structural equation modeling* (3^a ed.). New York, NY: Routledge. doi: [10.4324/9780203851319](https://doi.org/10.4324/9780203851319)
- Sierra, J. C., Vallejo-Medina, P., Santos-Iglesias, P., & Lameiras-Fernández, M. (2012). Validación del Massachusetts General Hospital-Sexual Functioning Questionnaire (MGH-SFQ) en población española. *Atención Primaria*, 44(9), 516-524. doi: [10.1016/j.aprim.2012.02.004](https://doi.org/10.1016/j.aprim.2012.02.004)
- Sprecher, S. (2002). Sexual satisfaction in premarital relationships: Associations with satisfaction, love, commitment, and stability. *The Journal of Sex Research*, 39(3), 190-196. doi: [10.1080/00224490209552141](https://doi.org/10.1080/00224490209552141)
- Sprecher, S., & Cate, R. M. (2004). Sexual satisfaction and sexual expression as predictors of relationship satisfaction and stability. En J. H. Harvey, A. Wenzel, S. Sprecher, J. H. Harvey, A. Wenzel & S. Sprecher (Eds.), *The Handbook of Sexuality in Close Relationships* (pp. 235-256). Mahwah, NJ: Lawrence Erlbaum.
- StataCorp. (2017). Stata: Release 15. [software de cómputo]. College Station, TX: StataCorp. LLC.
- Stephenson, K. R., Ahrold, T. K., & Meston, C. M. (2011). The association between sexual motives and sexual satisfaction: Gender differences and categorical comparisons. *Archives of Sexual Behavior*, 40(3), 607-618. doi: [10.1007/s10508-010-9674-4](https://doi.org/10.1007/s10508-010-9674-4)
- Stevens, J. (2002). *Applied multivariate statistics for the social sciences* (4^a ed.). Mahwah, NJ: Lawrence Erlbaum.
- Trompeter, S. E., Bettencourt, R., & Barrett-Connor, E. (2012). Sexual activity and satisfaction in healthy community-dwelling older women. *The American Journal of Medicine*, 125(1), 37-43. doi: [10.1016/j.amjmed.2011.11.011](https://doi.org/10.1016/j.amjmed.2011.11.011)

amjmed.2011.07.036

- Vizcarrondo-Godreau, M. C. (2001). *Construcción y normalización de un inventario de satisfacción sexual dirigido a mujeres* (Tesis doctoral inédita). Universidad Carlos Albizu, San Juan, Puerto Rico.
- World Health Organization. (2010). *Measuring sexual health: Conceptual and practical considerations and related indicators*. Recuperado de <http://apps.who.int>
- Yamamiya, Y., Cash, T. F., & Thompson, J. K. (2006). Sexual experiences among college women: The differential effects of general versus contextual body images on sexuality. *Sex Roles: A Journal of Research*, 55(5-6), 421-427. doi: [10.1007/s11199-006-9096-x](https://doi.org/10.1007/s11199-006-9096-x)
- Young, M., Denny, G., Luquis, R., & Young, T. (1998). Correlates of sexual satisfaction in marriage. *The Canadian Journal of Human Sexuality*, 7(2), 115-127.
- Young, M., Denny, G., Young, T., & Luquis, R. (2000). Sexual satisfaction among married women. *American Journal of Health Studies*, 16(2), 73-84. Disponible en <http://www.va-ajhs.com>
-

Actitud sobre el ejercicio físico y los deportes: Un estudio psicométrico en estudiantes universitarios

Attitudes towards physical exercise and sports: A psychometric study in college students

Andrés Cruz-Santos *¹, Juan A. González-Rivera², Adam Rosario-Rodríguez¹

1 - Universidad Carlos Albizu, San Juan, Puerto Rico.

2 - Ponce Health Sciences University, San Juan University Center, Puerto Rico.

Introducción
Método
Resultados
Discusión
Conclusión
Referencias

Recibido: 29/01/2019 Revisado: 06/03/2019 Aceptado: 31/03/2019

Resumen

El presente estudio analiza las propiedades psicométricas de la Escala de Actitud Hacia el Ejercicio Físico y los Deportes (AHEFD-11) en una muestra de estudiantes adultos puertorriqueños. Un total de 255 puertorriqueños participaron en este estudio de carácter exploratorio y psicométrico. Los resultados reflejaron una solución de tres factores como la más apropiada para explicar la varianza en las puntuaciones. Estos factores fueron identificados como *actitud hacia el ejercicio físico, pasión por el deporte y disposición*. Un total de 11 reactivos cumplieron con los criterios de discriminación y obtuvieron cargas factoriales apropiadas. El índice de consistencia interna alfa de Cronbach de la escala fue de .88. Además, se encontró una correlación positiva y estadísticamente significativa entre la actitud hacia el ejercicio físico y el deporte, el bienestar psicológico y la satisfacción con la vida ($r = .27, p < .001$). AHEFD-11 permitirá nuevas investigaciones sobre el fenómeno del ejercicio y el deporte en Puerto Rico.

Palabras clave: *actividad física, propiedades psicométricas, validación, ejercicio, deportes*

Abstract

The following study analyses the psychometric properties of the Attitudes Towards Physical Exercise and Sport Scale (AHEFD-11) in a sample of Puerto Rican adults. A total of 255 Puerto Ricans participated in this exploratory and psychometric study. The results reflected a three-factor solution as the most appropriate to explain the variance in the scores. These factors were identified as *attitude to physical exercise, passion for sport and disposition*. A total of 11 items met the discrimination criteria and obtained appropriate factor loadings. The reliability index (Cronbach's alpha) of the scale was .88. In addition, a positive and statistically significant correlation was found between attitude towards physical exercise and sports, psychological well-being and life satisfaction ($r = .27, p < .001$). The AHEFD-11 will allow new research on the phenomenon of physical exercise and sports in Puerto Rico.

Keywords: *physical activity, psychometric properties, validation, exercise, sports*

* Correspondencia a: Andrés Cruz Santos; Po Box 834 Cidra P.R 00739. Tel.: 787- 436-5950. Email: andcruz@albizu.edu

Cómo citar este artículo: Cruz-Santos, A., González-Rivera, J. A., & Rosario-Rodríguez, A. (2019). Actitud sobre el ejercicio físico y los deportes: Un estudio psicométrico en estudiantes universitarios. *Revista Evaluar*, 19(2), 58-72. Recuperado de <https://revistas.unc.edu.ar/index.php/revaluar>

Introducción

En las últimas décadas, se ha experimentado un aumento en la producción de nuevas tecnologías que facilitan las responsabilidades y quehaceres cotidianos. Cada día se observa cómo estos avances tecnológicos llegan al mercado para de alguna manera servir como extensión del cuerpo humano, ayudando a las personas a llevar a cabo actividades de manera simple y con menos esfuerzo físico. No obstante, estos avances han hecho que las personas disminuyan la actividad física durante sus vidas, aún si reconocen que el ejercicio físico está vinculado con beneficios psicológicos (Hassmen, Koivula, & Uutela, 2000).

La Organización Mundial de la Salud (World Health Organization [WHO], 2010) expresa que la inactividad física es un problema de salud pública a nivel mundial. Esto provoca un sinnúmero de enfermedades que han afectado a la población de manera negativa en estos últimos años. La causa principal del sobrepeso es la inactividad física y el desequilibrio energético entre las calorías consumidas y las gastadas. Al menos un 60% de la población mundial no realiza la actividad física necesaria para obtener beneficios para la salud (WHO, 2010). También, la WHO (2010) expone como consecuencia de la inactividad física a enfermedades como: cardiopatías coronarias, accidentes cardiovasculares, diabetes, hipertensión, distintos tipos de cáncer, como el cáncer del colon y de mama, y trastornos mentales como la depresión. El ejercicio físico habitual proporciona múltiples beneficios para la salud del individuo que lo realiza, dentro de estos, disminuir las probabilidades de mortalidad por causa de enfermedades crónicas cardiovasculares, metabólicas y oncológicas (Rebolledo-Cobos, Teixeira, & Correa, 2014). Meseguer-Zafra et al. (2019) mencionan que en la actualidad se están implementando diferentes programas de ejercicio físico como alternativa de

tratamiento. De esta manera, la actividad física se presenta como un instrumento útil para reducir la mortalidad derivada de este tipo de enfermedades y, por tanto, se debe fomentar como parte activa en el tratamiento.

Por una parte, epidemiólogos han estudiado los comportamientos de evasión de la actividad física como parte activa en el tratamiento y han determinado que estos impactan negativamente en la salud (Pedišić, Dumuid, & Olds, 2017). En efecto, existe una inquietud sobre si las personas que no realizan actividad física pueden gozar de bienestar psicológico y estar satisfechos con la vida. A estos efectos, esta investigación fue dirigida al desarrollo y validación una escala para medir la actitud hacia el ejercicio físico y el deporte y, a su vez, auscultar si existe una correlación significativa entre el ejercicio, el bienestar psicológico y la satisfacción con la vida en una muestra de estudiantes universitarios puertorriqueños.

Ejercicio Físico

Barbany (2015) explica que el acto del ejercicio físico está compuesto por la acción de componentes de tipo motor: los músculos de tipo esquelético forman parte del proceso por el cual la energía química se convierte directamente en energía mecánica, lo que genera una fuerza que permite que los movimientos se lleven a cabo. Otra definición expone que el ejercicio físico regular, el cual hace referencia al esfuerzo físico prolongado, representa una medida directa del estado general de salud biológica, especialmente del sistema cardiovascular (Taylor, Buskirk, & Henschel, 1955). En cambio, en el ámbito de la psicología de la actividad física y el deporte, se lo ha vinculado en los últimos años con un concepto llamado *fortaleza mental*, definido como

“un conjunto de valores, actitudes, emociones y cogniciones que influyen en la manera como una persona accede, responde y evalúa eventos de alto nivel de exigencia para lograr sus metas de manera consistente” (Gucciardi, Gordon, & Dimmock, 2009). También se menciona que el ejercicio físico permite a las personas llevar a cabo actividades cotidianas sin fatigarse (Malina & Katzmarzyk, 2006). A diferencia del ejercicio físico, la actividad física como caminar más rápido y aumentar la intensidad del trabajo, puede ayudar a cambiar el estilo de vida (Haith-Cooper, Waskett, Montague, & Horne, 2018). Según el *American College of Sports Medicine* (2000), los términos *ejercicio físico* y *actividad física* sostienen una similitud importante entre ellos. Sin embargo, también presentan diferencias. El ejercicio físico corresponde a un acto planificado y organizado de actividad física, mientras que la actividad física solo define el movimiento corporal que se puede llevar a cabo de distintas maneras (Hassmén et al., 2000). Otra diferencia entre el ejercicio físico y la actividad física es que la actividad física es un estado de condición que le permite a un individuo realizar actividades cotidianas sin sufrir fatiga y disfrutar adecuadamente de sus objetivos diarios. En este sentido, la actividad física se observa de manera histórica en tres componentes: fuerza muscular y resistencia, resistencia cardiorrespiratoria y habilidad motora.

La actividad física, a diferencia del ejercicio físico, no tiene que ser programada y organizada (Latorre-Román et al., 2017). Del mismo modo, el término *bienestar físico* está compuesto por varias características como aptitud respiratoria y composición corporal, que incluye a su vez distribución de grasa, fuerza muscular, flexibilidad y resistencia (Pollock et al., 1998). No obstante, Malina y Katzmarzyk (2006) explican la actividad física como una conducta con distintas dimensiones que conlleva el movimiento del

cuerpo a través del espacio. La actividad física se observa a menudo como un gasto de energía y estrés asociada a la pérdida de peso y la sensación de bienestar. Además, se observa como un componente específico de las destrezas de la aptitud física. El contexto sociocultural también define el tipo de actividad física que se lleva a cabo, como por ejemplo deportes, juego, educación, trabajo y ejercicio, entre otros. Estas actividades están fuertemente influenciadas por la cultura (Latorre-Román et al., 2017). Una buena forma de reconocer que se está realizando una actividad física adecuada es cuando el ritmo cardíaco aumenta, pero todavía se puede hablar (WHO, 2010).

En un estudio llevado a cabo en Inglaterra, Escocia, Gales e Irlanda del Norte se presentan algunas reglas mínimas sobre el volumen, la duración, la frecuencia y el tipo de actividad física necesarios durante toda la vida con el fin de lograr beneficios generales para el bienestar. Este estudio promueve también la actividad física, el deporte, el ejercicio físico, y los métodos de transportación activos como el caminar para lograr un bienestar adecuado (Department of Health, Physical Activity, Health Improvement and Protection, 2011). Además, Franco et al. (2015) mencionan unas estrategias para aumentar la práctica de ejercicio físico entre las personas que incluyen: 1) Aumentar la conciencia de los beneficios y minimizar los riesgos percibidos de la actividad física. En ocasiones, las personas desconocen los beneficios que puede brindar el ejercicio físico. 2) Mejorar el acceso ambiental y económico a las oportunidades del ejercicio físico. La accesibilidad de un lugar donde se puedan llevar a cabo ejercicios a bajo costo es mínima en el occidente (WHO, 2010). Jiménez, Martínez, Miró y Sánchez (2008) argumentan que la prescripción de ejercicio físico orientado a obtener determinados beneficios para la salud se relaciona frecuentemente con diferentes parámetros (frecuencia, duración, intensidad,

tipo de ejercicio, etc.). Por otra parte, [Hassmén et al. \(2000\)](#) comparten algunas recomendaciones para incluir el ejercicio físico en la vida de las personas y promover y mantener un buen estado de salud. Los adultos de 18 a 65 años necesitan 30 minutos de actividad física por al menos cinco días a la semana o 20 minutos de actividad física de alta intensidad al menos tres días a la semana. La combinación de ejercicio moderado y de alta intensidad puede ser utilizada para alcanzar el mínimo recomendado de actividad física por semana.

[Pollock et al. \(1998\)](#) definen los requisitos que se deben cumplir para lograr una capacidad cardiorrespiratoria y una composición corporal adecuada mediante el ejercicio físico, como, por ejemplo, la frecuencia y la planificación del entrenamiento: este debe ser llevado a cabo por un mínimo de tres días semanales y un máximo de cinco días semanales. La intensidad del entrenamiento debe ser de un 55/65% - 90% del ritmo cardíaco máximo de cada persona (HRmax), o 40/50% - 85% del consumo de reserva de oxígeno (V.O2R). La duración del entrenamiento debe ser de 20 a 60 minutos de actividad aeróbica continua o intermitente. En cuanto al modo o la actividad, se recomienda cualquier ejercicio que use grupos grandes de músculos, que se puedan trabajar continuamente como caminar-excursionismo, correr-trotar, ciclismo-montar la bicicleta, baile, remar, subir escaleras, nadar, patinar y varios deportes de alto rendimiento ([Pollock et al., 1998](#)). Por otro lado, el término *actitud* procede de la psicología social. Son muchas las definiciones que se le han dado a lo largo del siglo XX desde que nació el término como un constructo para explicar y medir el grado por el que las personas tienden a comportarse de manera selectivamente distinta en situaciones similares. La variación de una a otra definición es producto de la orientación que se le dé a la misma. De esta manera, algu-

nos autores dan prioridad a elementos cognitivos, otros a elementos afectivos y otros a elementos comportamentales ([Dosil-Díaz, 2002](#)).

Instrumentos para evaluar la actitud hacia el ejercicio físico y el deporte

La mayoría de los instrumentos con que se cuenta en nuestro ámbito han sido desarrollados en países sudamericanos y no toman en consideración la relación con el bienestar psicológico y la satisfacción con la vida. En los últimos años, los estudios sobre actitudes y la actividad física y el deporte han aumentado considerablemente ([Dosil-Díaz, 2002](#)). Los trabajos que más relevancia han tenido se deben a [Kenyon \(1968\)](#) quien desarrolló una línea de investigación sobre el tema que se cristalizó con la construcción de la Escala de Actitudes hacia la Actividad Física (ATPA). De manera análoga, uno de los instrumentos utilizados en investigaciones en Sudamérica es la Escala de Actitudes hacia la Actividad Física y el Deporte ([Dosil-Díaz, 2002](#)). Esta escala pretende ser un instrumento de medida que valora la actitud de cualquier sujeto hacia la actividad física y el deporte, independientemente de si es o no practicante. [Dosil-Díaz \(2002\)](#) recomienda que próximas investigaciones profundicen en el constructo *actitud*, teniendo en cuenta la relación, en principio positiva, entre la práctica y la actitud, pero sin asumir que se tiene una actitud negativa hacia el deporte solo porque no se practica ejercicio. A estos efectos, esta investigación tiene como objetivo el desarrollo y la validación de una escala que mida la actitud hacia el ejercicio físico y el deporte y, a su vez, examinar la relación de estos con la satisfacción con la vida y el bienestar psicológico y auscultar si existen diferencias estadísticamente significativas entre las medias de actitud hacia el ejercicio físico y el deporte entre

hombres y mujeres. Los investigadores establecieron como hipótesis que existen relaciones significativas entre la actitud hacia el ejercicio físico y los deportes y la satisfacción con la vida y el bienestar psicológico. Además, que existen diferencias significativas en cuanto a la actitud hacia el ejercicio físico y los deportes en relación con el género.

Método

Diseño de Investigación

El diseño que se utilizó en esta investigación fue ex post facto, de tipo instrumental (Montero & León, 2007).

Participantes

El procedimiento para la selección de los participantes fue por disponibilidad, de tipo no probabilístico. La muestra estuvo compuesta por 255 estudiantes universitarios adultos reclutados a través de correo electrónico y redes sociales. En la Tabla 1 se presentan los datos sociodemográficos de los participantes. Para participar de este estudio, se establecieron los siguientes criterios de inclusión: (1) ser mayor de 21 años de edad, (2) tener la capacidad de leer en español, (3) ser puertorriqueños, (4) ser estudiantes activos de una universidad.

Instrumentos

Cuestionario de Datos Generales. Desarrollamos un cuestionario ad-hoc de datos sociodemográficos para obtener información relacionada con la edad, sexo, estado civil, preparación académica, ingresos anuales, frecuencia de actividad física,

entre otras variables que permitieron la descripción de la muestra del estudio.

Tabla 1
Información sociodemográfica de la muestra.

Variables	f	%
Edad		
21 - 30	159	62.4
31 - 40	62	24.3
42 - 50	25	9.8
51 - 65	9	3.5
Sexo		
Femenino	203	79.6
Masculino	52	20.4
Preparación Académica		
Escuela superior o menos	10	3.9
Bachillerato	96	37.6
Maestría	109	42.7
Doctorado	40	15.7
Estado civil		
Casado/Convivencia	79	31.0
Soltero	162	63.5
Divorciado	12	4.7
Otros	2	0.8
Ingresos Anuales		
\$0 - \$25,000	169	66.3
\$26,000 - \$50,000	55	21.6
\$51,000 - \$100,00	22	8.6
\$101,000 o más	9	3.5
Frecuencia de actividad física		
Diariamente	33	12.9
Tres veces por semana	75	29.9
Dos veces por semana	35	13.7
Una vez por semana	33	12.9

Nota. N = 255.

Escala de Actitudes Hacia el Ejercicio Físico y el Deporte (AHEFD-11). Gracias a los trabajos de Dosil-Díaz (2002) y la construcción de la Escala de Actitudes hacia la Actividad Física y el Deporte en Sudamérica. A pesar de que los términos son parecidos, a diferencia de la escala de Dosil-Díaz (2002), que solo trabaja el concepto de actividad física, la escala AHEFD-11 también aborda conceptos más centrados en el ejercicio físico. A pesar de que los conceptos actividad

física y ejercicio físico son similares, existen diferencias significativas: a diferencia del ejercicio físico, que conlleva una planificación (Pollock et al., 1998), la actividad física no necesariamente es planificada o estructurada. Los autores de esta investigación desarrollaron 13 reactivos que se agruparon en tres subescalas: *actitud hacia el ejercicio, pasión por el deporte y disposición*. Estas dimensiones integran las actitudes, pasión y práctica de actividad física y deporte que puede tener un individuo. El análisis factorial exploratorio confirmó la presencia de estas tres dimensiones. En cuanto a las subescalas del instrumento, la dimensión denominada actitud hacia el ejercicio está asociada a la actitud que poseen las personas en cuanto a sus prioridades, consideraciones, tiempo y dedicación que muestran cuando realizan algún ejercicio físico. Estas características son importantes para el resultado de una actividad física adecuada (Pollock et al., 1998).

La escala fue desarrollada por el autor principal para medir la actitud hacia la actividad física y el deporte desde una perspectiva multidimensional. La versión original de la escala estaba constituida por 13 reactivos. La creación de los reactivos fue inspirada por los estudios de Pollock et al. (1998) y la WHO (2010). Estos reactivos fueron agrupados dentro de una estructura de tres factores según la taxonomía del creador de la escala: actitud hacia el ejercicio físico, pasión por el deporte y disposición. Se utilizó un formato de cuatro respuestas estilo Likert que fluctuaban desde *Totalmente en desacuerdo* hasta *Totalmente en acuerdo* en continuo numérico del 1 al 4. El puntaje más bajo que se podía obtener en la versión original es 13 y el más alto es 52, donde una mayor puntuación indica una actitud más positiva hacia el ejercicio físico.

Escala Breve de Bienestar Psicológico. Este instrumento fue desarrollado por Ryff (1989) y

adaptado a la población puertorriqueña por González-Rivera, Quintero-Jiménez, Veray-Alicea y Rosario-Rodríguez (2016). Se compone de 17 reactivos organizados en una escala tipo Likert de seis puntos que va de *Fuertemente en desacuerdo* a *De acuerdo* (por ejemplo; *No tengo claro lo que quiero conseguir en la vida*). La escala consta de cuatro factores: 1) dominio del entorno, 2) propósito de vida, 3) auto-aceptación, y 4) autonomía, y tiene un índice de confiabilidad alfa de Cronbach de .85.

Escala de Satisfacción con la Vida. Esta escala es la versión en español del instrumento Satisfaction With Life Scale (SWLS) de Diener, Emmons, Larsen y Griffin (1985). Estos autores y la WHO (2010) consideran que la satisfacción con la vida constituye el componente cognitivo del bienestar subjetivo. El instrumento está constituido por un total de 5 reactivos organizados en una escala tipo Likert de siete puntos que va de *Muy en desacuerdo* a *Muy de acuerdo* (por ejemplo, *Soy bastante bueno manejando las muchas responsabilidades de mi vida cotidiana; Estoy satisfecho con mi vida; Si pudiera vivir mi vida de nuevo, no cambiaría nada*). Este instrumento tiene un índice de consistencia interna alfa de Cronbach de .88.

Procedimientos Generales

Los datos fueron recopilados de forma electrónica durante el mes de diciembre 2018. Se distribuyó un anuncio a través de redes sociales (Facebook) y correos electrónicos para que las personas interesadas pudieran participar de la investigación. Los participantes tenían autorización para compartir el anuncio en sus redes sociales y por correo electrónico, con el fin de ampliar la muestra mediante el efecto “bola de nieve”. Para la versión electrónica, se utilizó la plataforma de

PsychData (s. f.). Este es un sitio web dirigido a crear instrumentos en línea para el campo de la psicología que permitió que el instrumento fuera colocado en internet con todas las especificaciones del Comité de Ética en la Investigación de la Universidad Carlos Albizu recinto de San Juan.

Una vez recopilados los datos, analizamos los mismos utilizando el sistema para análisis estadísticos Statistical Package for the Social Sciences (IBM Corporation, 2016). Específicamente, realizamos análisis descriptivos de la muestra, análisis de reactivos para conocer su índice de discriminación, análisis de factores utilizando el método de extracción máxima verosimilitud con rotación oblicua, análisis de confiabilidad para calcular el coeficiente alfa de Cronbach y división en mitades de Spearman-Brown, análisis de comparación y análisis de correlación. Todos los procedimientos de este estudio fueron aprobados por el Comité Institucional de Revisión (IRB por sus siglas en inglés) de la Universidad Carlos Albizu de San Juan, Puerto Rico.

Resultados

Análisis factorial exploratorio

Para determinar la validez de construcción lógica del instrumento, llevamos a cabo un análisis factorial utilizando el método de extracción de máxima verosimilitud con rotación oblicua para factores correlacionados, identificando aquellos factores que expliquen el 5% o más de la varianza, como sugiere Hatcher (1994). Como criterios de aceptación, consideramos aquellos reactivos con una carga factorial mayor a .50 en un solo factor (Stevens, 2002). Los resultados mostraron una estructura de tres factores que explicaban el 69% de la varianza de los datos originales. Sin embargo, cuando revisamos las cargas factoriales de cada reactivo, se identificaron dos ítems que

no cumplían con los criterios de aceptación detallados anteriormente y que fueron eliminados de los análisis posteriores: (3) *Alguna vez me he acostado o levantado antes de mi horario acostado para poder llevar a cabo un ejercicio físico o deportivo*, y (10) *Si dejo de hacer ejercicio físico o deporte por una semana, observo o experimento una reducción en mi condición física*. Luego, procedimos a realizar nuevamente el análisis factorial con los 11 reactivos restantes.

En esta segunda ocasión, los resultados mostraron nuevamente una estructura de tres factores que explicaban el 69% de la varianza, del cual el 46% es explicado por el Factor 1, el 13% es explicado por el Factor 2 y el 10% es explicado por el Factor 3. Al revisar las cargas factoriales de estos reactivos, observamos que todos cumplieron los criterios de inclusión. Las pruebas de Kaiser-Meyer-Olkin ($KMO = .890$) y la prueba de esfericidad de Bartlett ($\chi^2_{(55)} = 1297.078, p < .001$) confirmaron la adecuación de los datos de muestreo para el análisis. En la Tabla 2 se presentan las cargas factoriales obtenidas por los 11 reactivos, cuya distribución por dimensión fue la siguiente: 5 reactivos en el primer factor, 3 en el segundo y 3 en el tercero.

Al revisar detenidamente la agrupación de los reactivos por su carga factorial, se determinó que los reactivos 11, 1, 13, 7 y 8 pertenecen a la dimensión de la actitud de la persona hacia el ejercicio físico. Los reactivos 6, 12 y 4 pertenecen a la dimensión de la pasión por el deporte que muestran las personas. Finalmente, los reactivos 5, 9 y 2 pertenecen a la dimensión de la predisposición de la persona para realizar actividad física.

Análisis de reactivos

Se analizó la discriminación de los 11 ítems de la AHEFD-11 en sus respectivos factores y en

Tabla 2

Reactivos cuyas cargas factoriales fueron superiores a .50 y sus factores correspondientes.

Reactivo	1	2	3
11. Entre mis prioridades en la vida está el ejercicio físico y deportivo.	.890		
1. El ejercicio físico y deportivo ocupa un lugar importante en mi vida.	.872		
13. En general, me considero una persona que tiene una actitud favorable o de buena predisposición hacia el ejercicio físico o deportivo.	.685		
7. Le dedico más tiempo a la práctica de algún ejercicio físico o deportivo que a pasar tiempo con mis amistades.	.587		
8. Mis deseos de realizar o practicar ejercicios físicos no disminuyen aún cuando no pueda practicarlos durante una semana completa.	.559		
6. Alguna vez me he acostado o levantado antes de mi horario acostumbrado para poder ver una retransmisión deportiva.		.702	
12. Mis programas favoritos son los deportivos.		.684	
4. Siempre que tengo la oportunidad asisto a manifestaciones o espectáculos deportivos presencialmente (juegos de pelota, baloncesto, voleibol, entre otras).		.628	
5. Me gustaría tener siempre un tiempo en la semana para dedicarlo al ejercicio físico o deporte.			.756
9. Mis deseos de realizar o practicar ejercicio físico aumentan cuando veo a una persona practicándolo.			.649
2. Si tuviera una tarde libre, emplearía una parte de mi tiempo en realizar algún ejercicio físico o deportivo.			.639

Nota. Factor 1 = Actitud hacia el ejercicio físico; Factor 2 = Pasión por el deporte; Factor 3 = Disposición.

la escala total mediante el índice de correlación ítem-total. Los índices de discriminación de los ítems para la dimensión actitud hacia el ejercicio físico fluctuaron entre .56 y .81, mientras que para la pasión deportiva oscilaron entre .56 y .57. Asimismo, la puntuación del factor disposición osciló entre .57 y .64 (véase Tabla 3). Todos los índices de discriminación de la versión trifactorial de la AHEFD-11 sobrepasaron el valor mínimo recomendado de .30 (Field, 2013).

Análisis de confiabilidad

Los 11 reactivos fueron sometidos a un análisis de consistencia interna para determinar el índice de confiabilidad de la escala. Para ello utilizamos dos métodos: el coeficiente alfa de Cronbach y la división en mitades de Spearman-Brown. Los resultados mostraron un coeficiente alfa de .72 y

de .90 para la prueba Spearman-Brown. Utilizamos estos mismos análisis para cada una de las dimensiones identificadas en el análisis de factores. En la Tabla 4 se presentan el coeficiente alfa de Cronbach, el coeficiente de la prueba Spearman-Brown, la media y la desviación estándar de la versión final de la AHEFD-11 y sus subescalas.

Análisis de correlación

Se realizó un análisis correlacional lineal de los tres factores identificados y la AHEFD-11, la escala de bienestar psicológico y la escala de satisfacción con la vida utilizando el coeficiente producto-momento de Pearson. Los resultados se muestran en la Tabla 4. Como se puede observar, todos los coeficientes r de Pearson obtenidos en los análisis fueron estadísticamente significativos.

Tabla 3

Índices de discriminación y varianza explicada de los ítems en sus tres dimensiones (actitud hacia el ejercicio físico, pasión por el deporte y disposición).

Ítems	r_{bis}^1	r_{bis}^2	r_{bis}^3	r_{bis}^T
11. Entre mis prioridades en la vida está el ejercicio físico y deportivo.	.81			.75
1. El ejercicio físico y deportivo ocupa un lugar importante en mi vida.	.81			.74
13. En general, me considero una persona que tiene una actitud favorable o de buena predisposición hacia el ejercicio físico o deportivo.	.73			.72
7. Le dedico más tiempo a la práctica de algún ejercicio físico o deportivo que a pasar tiempo con mis amistades.	.56			.53
8. Mis deseos de realizar o practicar ejercicio físico no disminuyen, aún cuando no pueda practicarlos durante una semana completa.	.63			.62
6. Alguna vez me he acostado o levantado antes de mi horario acostumbrado para poder ver una retransmisión deportiva.		.56		.43
12. Mis programas favoritos son los deportivos.		.56		.51
4. Siempre que tengo la oportunidad, asisto a manifestaciones o espectáculos deportivos presencialmente (juegos de pelota, baloncesto, voleibol, entre otros).		.56		.46
5. Me gustaría tener siempre un tiempo en la semana para dedicarlo al ejercicio físico o deporte.			.64	.52
9. Mis deseos de realizar o practicar ejercicio físico aumentan cuando veo a una persona practicándolo.			.57	.47
2. Si tuviera una tarde libre, emplearía una parte de mi tiempo en realizar algún ejercicio físico o deportivo.			.64	.62

Nota. r_{bis}^1 = índice de discriminación para actitud hacia el ejercicio físico y el deporte; r_{bis}^2 = índice de discriminación para pasión por el deporte; r_{bis}^3 = índice de discriminación para disposición; r_{bis}^T = índice de discriminación de la escala total.

Los resultados reflejaron que existe una correlación moderada baja tanto entre la actitud hacia el ejercicio físico y el bienestar psicológico como entre la actitud hacia el ejercicio físico y la satisfacción con la vida. Estos resultados respaldan la información provista por Pollock et al. (1998) y por Jiménez et al. (2008), quienes mencionan una correlación significativa entre el ejercicio físico y determinados efectos saludables, como el bienestar y la satisfacción con la vida.

Análisis de comparación

Se realizó un análisis de comparación de medias mediante prueba *t* para muestras independientes con el propósito de evaluar si existen

diferencias estadísticamente significativas entre las medias de actitud hacia el ejercicio físico y el deporte de hombres y mujeres (véase Tabla 5). Los análisis reflejaron que los hombres tienen una actitud más positiva hacia el ejercicio físico y el deporte que las mujeres. También se realizó una prueba *t* para evaluar si existen diferencias estadísticamente significativas entre las medias de bienestar psicológico y de satisfacción con la vida de las personas que practican ejercicio físico tres veces o más por semana y las medias de las que no. Los resultados revelaron que no existen diferencias significativas entre estos dos grupos.

Discusión

El presente estudio tuvo como objetivo prin-

Tabla 4

Correlaciones entre las variables, sus medias, desviaciones estándar y alfa de Cronbach.

	α	S-B	M	DE	1	2	3	4	5
1. Actitud hacia el ejercicio	.88	.87	12.47	3.97	-				
2. Pasión por el deporte	.73	.72	5.82	2.73	.92*	-			
3. Disposición	.77	.78	15.14	1.48	.71*	.46*	-		
4. AHEFD-1	.88	.90	27.78	6.88	.76*	.59*	.33*	-	
5. Bienestar psicológico	.85	.88	80.80	11.19	.34*	.31*	.22*	.29*	-
6. Satisfacción con la vida	.88	.85	25.96	6.24	.27*	.26*	.20*	.17*	.74*

Nota. * = $p < .01$; α = alfa de Cronbach; S-B = Spearman-Brown; M = medias; DE = desviación estándar.

principal desarrollar y validar la Escala de Actitud Hacia el Ejercicio Físico y los Deportes (AHEFD-11) en una muestra de adultos universitarios puertorriqueños. Este es el primer instrumento desarrollado totalmente en Puerto Rico para medir la actitud hacia el ejercicio físico y los deportes desde una perspectiva holística y contextualizada que toma en consideración la historia y la cultura de los puertorriqueños. Además, este estudio pretendía conocer la relación que existe entre la actitud hacia el ejercicio físico y los deportes, y entre la satisfacción con la vida y el bienestar psicológico. Esto representa un importante aporte para el desarrollo de nuevas investigaciones sobre

la actitud hacia el ejercicio físico y los deportes en Puerto Rico. A su vez, permitirá análisis más amplios sobre las diferencias entre aspectos, tales como la actitud hacia la actividad física (practicada o no), la satisfacción con la vida, el bienestar psicológico y el impacto de estos en la salud física y psicológica de los puertorriqueños.

Los fundamentos teóricos y prácticos de [Dósil-Díaz \(2002\)](#) permitieron desarrollar 13 reactivos que se agruparon en tres subescalas: actitud hacia el ejercicio, pasión por el deporte y disposición. Estas dimensiones integran las actitudes, la pasión y las prácticas referentes a la actividad física y los deportes que puede manifestar un indi-

Tabla 5Resultados de la prueba t para el análisis de comparación de medias y tamaño del efecto.

Variables	n	M	DE	t	df	p	d
Actitud hacia el ejercicio							
Femenino	203	27.36	6.72	-2.20	252	.03	.34
Masculino	51	29.71	7.12				
Bienestar psicológico							
Tres veces por semana	108	83.03	10.09	1.36	174	.18	.21
Menos de tres veces	68	80.87	10.53				
Valores perdidos	79						
Satisfacción con la vida							
Femenino	108	26.71	6.45	0.29	174	.27	.06
Masculino	68	26.34	5.20				
Valores perdidos	79						

Nota. n = participantes; M = medias; df = grados de libertad; DE = desviación estándar; t = valor de la prueba t ; p = significancia; d = tamaño del efecto d de Cohen. (N = 255).

viduo. El análisis factorial exploratorio confirmó la presencia de estas tres dimensiones. En cuanto a las subescalas del instrumento, la dimensión denominada *actitud hacia el ejercicio* corresponde a la actitud de las personas hacia el ejercicio físico según sus prioridades, consideraciones, tiempo y dedicación (Pollock et al., 1998). Estas características son importantes para el resultado de una actividad física adecuada.

La segunda dimensión, *pasión por el deporte*, está asociada a los factores relacionales de la pasión deportiva que experimentan las personas por algún deporte en particular. Se trata de asistir, observar y modificar el estilo de vida para participar en algún evento deportivo, sea de manera presencial o como espectador a través de los diversos medios de comunicación. Según Dosil-Díaz (2002), existen elementos comportamentales y afectivos que describen esta dimensión y distinguen algunas características del apasionado por el deporte.

La tercera dimensión, *disposición*, está relacionada con los factores de disponibilidad de la persona para realizar algún ejercicio físico. Gucciardi et al. (2009) explican que, en el ámbito de la psicología de la actividad física y el deporte, se emplea un concepto llamado *fortaleza mental*, definido como “una colección de valores, actitudes, emociones y cogniciones que influyen en la manera como una persona accede, responde y evalúa eventos de alto nivel de exigencia para lograr sus metas de manera consistente”.

A estos efectos, la predisposición de la persona para realizar actividad física y deportiva puede ser una característica de la manera como el individuo evalúa eventos de alto nivel de exigencia. En cuanto a los hallazgos sobre la actitud hacia el ejercicio físico, el bienestar psicológico y la satisfacción con la vida, todos los coeficientes r de Pearson obtenidos en los análisis fueron estadísticamente significativos. Los resultados re-

flejaron que existe una correlación moderada baja tanto entre la actitud hacia el ejercicio físico y el bienestar psicológico como entre la actitud hacia el ejercicio físico y la satisfacción con la vida. Estos resultados concuerdan con De Miguel-Calvo, Schweiger-Gallo, de las Mozas-Majano y Hernández-López (2011) que comentan que realizar ejercicio físico de forma sistemática y controlada tiene una influencia positiva en la productividad y el bienestar de las personas. Del mismo modo, estos resultados también concuerdan con Jiménez-Moral, Zagalaz-Sánchez, Molero, Pulido-Martos y Ruiz (2013), quienes argumentan que existe una asociación positiva entre el ejercicio físico, la felicidad subjetiva y la satisfacción con la vida. En consecuencia, estos resultados contribuyen a entender más la importancia que tiene el ejercicio físico para las personas como práctica dirigida a mejorar su calidad de vida.

Por otro lado, se realizó un análisis de comparación de medias mediante prueba t con el propósito de evaluar la existencia de diferencias estadísticamente significativas entre las medias de actitud hacia el ejercicio físico y el deporte de hombres y mujeres. Se encontró que los hombres poseen una mejor actitud hacia el ejercicio físico que las mujeres. Estos resultados van alineados con Cossío-Bolaños et al. (2017), quien encontró en su estudio *Patrones de actividad física de adolescentes escolares: Validez, confiabilidad y propuesta de percentiles para su evaluación* que los hombres presentan mayores niveles de actitud física que las mujeres. También, Sevil-Serrano, Práxedes-Pizarro, Zaragoza-Casterad, del Villar-Álvarez y García-González (2017) en su estudio *Barreras percibidas para la práctica de actividad física en estudiantes universitarios: Diferencias por género y niveles de actividad física* mencionan que, atendiendo al género, los hombres obtienen mayor cumplimiento de práctica de actividad física (76.7%) y mayores niveles de

actividad física moderada y vigorosa semanales (384.93) que las mujeres. Esto puede explicar las diferencias significativas que existen entre hombres y mujeres en cuanto a la actividad física.

Como menciona la WHO (2010), en el año 2010 a nivel mundial, alrededor del 23% de los adultos mayores de 18 años no se mantenía suficientemente activo (un 20% de los hombres y un 27% de las mujeres). En los países de ingresos altos, el 26% de los hombres y el 35% de las mujeres no hacía suficiente ejercicio físico, frente a un 12% de los hombres y un 24% de las mujeres en los países de ingresos bajos. En consecuencia, se percibe un patrón de diferencias entre géneros en cuanto al ejercicio físico. Los roles creados por el entorno sociocultural a través del tiempo han generado una serie de conductas, valores y tareas diferenciadas por género. Las tareas menos onerosas estarían reservadas a la mujer y al hombre, todas las tareas físicas (Carrillo-Montoya, Tereso-Ramírez, & González-Ramírez, 2018). El contexto también define el tipo de actividad física que se lleva a cabo, por ejemplo, deportes, juego, educación, trabajo y ejercicio, entre otros. Estas distintas actividades están fuertemente influenciadas por la cultura (Latorre-Román et al., 2017). Esto puede explicar la causa por la que el hombre manifiesta, a lo largo del tiempo, una actitud más positiva hacia el ejercicio físico y los deportes que las mujeres.

Estos resultados no guardan relación con el trabajo de Jiménez et al. (2008) quienes mencionan que la prescripción de ejercicio físico va encaminada habitualmente a lograr determinados efectos saludables en relación con diferentes parámetros (frecuencia, duración intensidad, tipo de ejercicio, etc.). En el estudio en cuestión, las personas que realizan ejercicio físico dos días o menos a la semana no lo hacen de acuerdo con la frecuencia, duración e intensidad recomendadas por la WHO (2018). Sin embargo, en los resultados

obtenidos en esta investigación, los participantes no presentaron diferencias significativas en cuanto a su bienestar psicológico y satisfacción con la vida. Por lo tanto, estos resultados no están en la misma línea de otras investigaciones.

Tomando esto en consideración, dentro de la gran cantidad de estudios dirigidos a conocer los beneficios del ejercicio físico, se ha demostrado que el ejercicio físico tiene un efecto beneficioso para la salud (Barbany, 2015; Jiménez-Moral et al., 2013). Por otra parte, otros autores argumentan que, pese a los beneficios conocidos sobre la actividad física, un 30% de la población mundial no alcanza los niveles de actividad física recomendados (Franco et al., 2015). En cuanto a los efectos de bienestar psicológico y satisfacción con la vida, las personas que realizan actividad física con mayor frecuencia muestran efectos negativos menores en la salud y el bienestar psicológico (Haith-Cooper et al., 2018). Por el contrario, en el presente estudio, quienes realizan ejercicio físico dos veces o menos por semana no mostraron diferencias significativas con respecto al bienestar psicológico y la satisfacción con la vida. Franco et al. (2015) pueden brindar una posible explicación de este fenómeno. Franco menciona que, a pesar de que las personas manifiestan tener alguna idea sobre los beneficios del ejercicio y los deportes, solo un 30% practica alguna actividad física. Sin embargo, entre las causas de este bajo porcentaje se encuentran la baja conciencia de los beneficios que puede traer la actividad física y el escaso acceso a las condiciones ambientales y económicas que permiten su práctica. Esto, puede explicar por qué no hubo diferencias significativas entre las personas que llevaban a cabo ejercicios físicos tres veces o más por semana y las que no con respecto a sus niveles de satisfacción con la vida y bienestar psicológico.

En síntesis, las propiedades psicométricas obtenidas por la AHEFD-11 demuestran ín-

dices de confiabilidad, de consistencia interna y de validez de acuerdo con lo esperado. Esto fue verificado al observar la consistencia interna del instrumento final. De igual forma, la validez de construcción lógica del instrumento fue examinada a través de un análisis factorial exploratorio. Los análisis psicométricos de la AHEFD-11 demuestran que el instrumento es apropiado. Sin embargo, al ser un instrumento desarrollado en un estudio exploratorio, se lo debe poner a prueba utilizando una muestra mayor. Finalmente, esta escala podrá ser utilizada como una herramienta útil para aquellos profesionales que deseen auscultar las actitudes de las personas hacia el ejercicio físico y los deportes. Del mismo modo, detectar a individuos que necesiten psicoeducación en este aspecto, utilizando la escala como una herramienta para la prevención de enfermedades por falta de actividad física y, a su vez, para la promoción del ejercicio físico y los deportes.

Limitaciones

El tipo de muestreo utilizado fue por disponibilidad, lo que conlleva una reducida generalización de los resultados, que son relevantes sólo para los participantes del estudio. Por otra parte, dado el uso de plataformas electrónicas, se puede esperar que la muestra sea más representativa de un sector poblacional (usuarios frecuentes de tecnología). Por último, no se pudo examinar la confiabilidad de la escala a través del tiempo y sólo se examinó la confiabilidad de las escalas a través de sus elementos.

Recomendaciones

Recomendamos administrar la AHEFD-11 a otra muestra de participantes para realizar el

proceso de validación cruzada, así como para evaluar la validez convergente y divergente del instrumento final. Por último, recomendamos la validación de la AHEFD-11 en otras poblaciones hispanoamericanas para auscultar sus propiedades psicométricas.

Conclusión

La construcción y validación de la Escala de Actitud Hacia el Ejercicio Físico y los Deportes aportan al quehacer psicológico en la isla y al desarrollo de nuevos conocimientos sobre la actitud del puertorriqueño hacia la actividad física y los deportes. En primer lugar, la AHEFD-11 tiene la capacidad de contribuir a investigaciones científicas en el contexto sociocultural puertorriqueño que permitan entender con mayor profundidad la actitud de las personas hacia la actividad física y los deportes, tanto positiva como negativa, y su repercusión en la salud física y mental de la población. Por otra parte, en el ambiente clínico, la AHEFD-11 contribuye como herramienta práctica que los psicólogos pueden incorporar para tratar temas relacionados con las actitudes hacia el ejercicio, la actividad física, la salud física y los deportes como medio de distracción, entretenimiento y estilo de vida. Las tres subescalas del instrumento permiten al terapeuta identificar la actitud del cliente hacia el ejercicio o la participación en algún evento deportivo como método de distracción. No cabe duda de que estos dos aportes incrementan el valor de este instrumento en Puerto Rico.

Referencias

American College of Sports Medicine. (2000). *ACSMs guidelines for exercise testing and prescription*

- (6^a ed.). Philadelphia, PA: Lippincott Williams & Wilkins.
- Barbany, J. R. (2015). *Fisiología del ejercicio físico y del entrenamiento*. Barcelona, España: Paidotribo.
- Carrillo-Montoya, T. J., Tereso-Ramírez, L., & González-Ramírez, E. (2018). La familia y la construcción del posicionamiento laboral en las mujeres. Un análisis desde el Trabajo Social. *Prospectiva, Revista de Trabajo Social e Intervención Social*, 25, 91-111. doi: [10.25100/prts.v0i25.5802](https://doi.org/10.25100/prts.v0i25.5802)
- Cossío-Bolaños, M., Méndez-Cornejo, J., Luarte-Rocha, C., Vargas-Vitoria, R., Canqui-Flores, B., & Gomez-Campos, R. (2017). Patrones de actividad física de adolescentes escolares: Validez, confiabilidad y propuesta de percentiles para su evaluación. *Revista Chilena de Pediatría*, 88(1), 73-82. doi: [10.1016/j.rchipe.2016.07.010](https://doi.org/10.1016/j.rchipe.2016.07.010)
- De Miguel-Calvo, J. M., Schweiger-Gallo, I., de las Mozas-Majano, O., & Hernández-López, J. M. (2011). Efecto del ejercicio físico en la productividad laboral y el bienestar. *Revista de Psicología del Deporte*, 20(2), 589-604. Recuperado de <https://www.rpd-online.com>
- Department of Health, Physical Activity, Health Improvement and Protection. (2011). *Start Active, Stay Active: A report on physical activity for health from the four home countries' chief medical officers*. Recuperado de <https://www.gov.uk>
- Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The Satisfaction With Life Scale. *Journal of Personality Assessment*, 49(1), 71-75. doi: [10.1207/s15327752jpa4901_13](https://doi.org/10.1207/s15327752jpa4901_13)
- Dosil-Díaz, J. (2002). Escala de Actitudes hacia la Actividad Física y el Deporte (E.A.F.D). *Cuadernos de Psicología del Deporte*, 2(2), 43-55. Recuperado de <https://revistas.um.es/cpd/index>
- Field, A. P. (2013). *Discovering statistics using SPSS* (4^a ed.). Londres, Inglaterra: SAGE.
- Franco, M. R., Tong, A., Howard, K., Sherrington, C., Ferreira, P. H., Pinto, R. Z., & Ferreira, M. L. (2015). Older people's perspectives on participation in physical activity: A systematic review and thematic synthesis of qualitative literature. *British Journal of Sports Medicine*, 49(19), 1268-1276. doi: [10.1136/bjsports-2014-094015](https://doi.org/10.1136/bjsports-2014-094015)
- González-Rivera, J. A., Quintero-Jiménez, N., Veray-Alicea, J., & Rosario-Rodríguez, A. (2016). Adaptación y validación de la Escala de Bienestar Psicológico de Ryff en una muestra de adultos puertorriqueños. *Salud y Conducta Humana*, 3(1), 1-14. Recuperado de <http://rsych.com>
- Gucciardi, D. F., Gordon, S., & Dimmock, J. A. (2009). Advancing mental toughness research and theory using personal construct psychology. *International Review of Sport and Exercise Psychology*, 2(1), 54-72. doi: [10.1080/17509840802705938](https://doi.org/10.1080/17509840802705938)
- Haith-Cooper, M., Waskett, C., Montague, J., & Horne, M. (2018). Exercise and physical activity in asylum seekers in Northern England: Using the theoretical domains framework to identify barriers and facilitators. *BMC Public Health*, 18(1), 1-11. doi: [10.1186/s12889-018-5692-2](https://doi.org/10.1186/s12889-018-5692-2)
- Hassmén, P., Koivula, N., & Uutela, A. (2000). Physical exercise and psychological well-being: A population study in Finland. *Preventive Medicine*, 30(1), 17-25. doi: [10.1006/pmed.1999.0597](https://doi.org/10.1006/pmed.1999.0597)
- Hatcher, L. (1994). *A step-by-step approach to using the SAS System for Factor Analysis and Structural Equation Modeling*. Cary, NC: SAS Institute.
- IBM Corporation. (2016). IBM SPSS Statistics for Windows, Version 24.0 [Software de cómputo]. Armonk, NY: IBM.
- Jiménez, M. G., Martínez, P., Miró, E., & Sánchez, A. I. (2008). Bienestar psicológico y hábitos saludables: ¿Están asociados a la práctica de ejercicio físico? *International Journal of Clinical Health & Psychology*, 8(1), 185-202. Recuperado de <http://www.aepc.es/ijchp/busca.php?coid=English?>
- Jiménez-Moral, J. A., Zagalaz-Sánchez, M. L., Molero, D., Pulido-Martos, M., & Ruiz, J. R. (2013). Capacidad aeróbica, felicidad y satisfacción con la vida en adolescentes españoles. *Revista de Psicología del De-*

- porte*, 22(2), 429-436. Recuperado de <https://www.rpd-online.com>
- Kenyon, G. S. (1968). Six scales for assessing attitude toward physical activity. *Research Quarterly. American Association for Health, Physical Education and Recreation*, 39(3), 566-574. doi: 10.1080/10671188.1968.10616581
- Latorre-Román, P. A., Moreno del Castillo, R., Lucena-Zurita, M., Salas-Sánchez, J., García-Pinillos, F., & Mora-López, D. (2017). Physical fitness in preschool children: Association with sex, age and weight status. *Child: Care, Health and Development*, 43(2), 267-273. doi: 10.1111/cch.12404
- Malina, R. M., & Katzmarzyk, P. T. (2006). Physical activity and fitness in an international growth standard for preadolescent and adolescent children. *Food and Nutrition Bulletin*, 27(4), 295-313. doi: 10.1177/15648265060274S511
- Meseguer-Zafra, M., Rosa-Guillamón, A., García-Cantó, E., Rodríguez-García, P., Pérez-Soto, J. J., Tárraga-López, P. J., ...Tárraga-Marcos, M. (2019). Influence of a programme of therapeutic exercise on different clinical indicators related to dyslipidaemia in adult subjects aged between 26 to 73 years with a cardiovascular risk factor. *Hipertensión y Riesgo Vascular*, 36(1), 21-27. doi: 10.1016/j.hipert.2018.02.001
- Montero, I., & Leon, O. G. (2007). A guide for naming studies in psychology. *International Journal of Clinical and Health Psychology*, 7(3), 847-862. Recuperado de <http://www.aepc.es/ijchp/busca.php?coid=English?>
- Pedišić, Ž., Dumuid, D., & Olds, T. S. (2017). Integrating sleep, sedentary behaviour, and physical activity research in the emerging field of time-use epidemiology: Definitions, concepts, statistical methods, theoretical framework, and future directions. *Kinesiology*, 49(2), 252-269. Recuperado de <https://hr-cak.srce.hr/ojs/index.php/kinesiology/index>
- Pollock, M. L., Gaesser, G. A., Butcher, J. D., Després, J. P., Dishman, R. K., Franklin, B. A., & Garber, C. E. (1998). ACSM position stand: The recommended quantity and quality of exercise for developing and maintaining cardiorespiratory and muscular fitness, and flexibility in healthy adults. *Medicine & Science in Sports & Exercise*, 30(6), 975-991. doi: 10.1249/00005768-199806000-00032
- PsychData. (s. f.). Create online surveys with confidence and ease [Herramienta en línea]. Recuperado de <https://www.psychdata.com>
- Rebolledo-Cobos, R. C., Teixeira, B., & Correa, C. (2014). Asma, estrés oxidativo y ejercicio físico: Revisión de la literatura. *Fisioterapia*, 36(1), 40-48. doi: 10.1016/j.ft.2013.06.002
- Ryff, C. D. (1989). Happiness is everything, or is it? Explorations on the meaning of psychological well-being. *Journal of Personality and Social Psychology*, 57(6), 1069-1081. doi: 10.1037/0022-3514.57.6.1069
- Sevil-Serrano, J., Práxedes-Pizarro, A., Zaragoza-Castead, J., del Villar-Álvarez, F., & García-González, L. (2017). Barreras percibidas para la práctica de actividad física en estudiantes universitarios: Diferencias por género y niveles de actividad física. *Universitas Psychologica*, 16(4), 1-15. doi: 10.11144/Javeriana.upsy16-4.bppa
- Stevens, J. (2002). *Applied multivariate statistics for the social sciences* (4ª ed.). Mahwah, NJ: Lawrence Erlbaum.
- Taylor, H. L., Buskirk, E., & Henschel, A. (1955). Maximal oxygen intake as an objective measure of cardio-respiratory performance. *Journal of Applied Physiology*, 8(1), 73-80. doi: 10.1152/jappl.1955.8.1.73
- World Health Organization. (2010). *Global recommendations on physical activity for health*. Suiza: World Health Organization. Recuperado de <https://www.who.int>
- World Health Organization. (2018). *Benefits of physical activity and risks of an insufficient level of physical activity*. Recuperado de <https://www.who.int>

Adaptación de la Children's Test Anxiety Scale (CTAS) para su empleo en adolescentes argentinos

Adaptation of the Children's Test Anxiety Scale (CTAS) for its use in Argentine adolescents

Introducción
Metodología
Resultados
Discusión
Conclusiones
Referencias

Natalia Belén Rivera-Flores¹, Luis Alberto Furlan^{* 1}, Gerardo Aníbal Gnani¹

1- Laboratorio de Evaluación Psicológica y Educativa, Facultad de Psicología, Universidad Nacional de Córdoba. Córdoba, Argentina.

Recibido: 16/03/2019 Revisado: 22/05/2019 Aceptado: 02/06/2019

Resumen

Se adaptó al español la Children's Test Anxiety Scale (CTAS; Wren & Benson, 2004) para su empleo en adolescentes argentinos de entre 12 y 17 años. La CTAS es un autoinforme con 30 ítems agrupados en 3 dimensiones: *comportamiento fuera de la tarea, pensamientos y reacciones autonómicas*. Se realizó la traducción directa de los ítems y el análisis de equivalencia en una muestra bilingüe. Mediante análisis factorial exploratorio (N = 360 estudiantes, de ambos sexos, de escuelas secundarias públicas y privadas, de 11 a 13 años, M = 12.77, DE = .72) se obtuvieron dos soluciones alternativas para su estructura, una con las tres dimensiones originales y otra de cuatro factores, diferenciando pensamientos de preocupación y autocríticos. Se evaluó el ajuste de ambos modelos con análisis factorial confirmatorio (N = 501 estudiantes, de ambos sexos, de escuelas secundarias públicas y privadas, de 12 a 17 años, M = 13.98, DE: .87), obteniendo el de cuatro dimensiones los índices más aceptables ($\chi^2/df = 1.92$, CFI = .90, TLI = .09, RMSEA = 0.043, AIC = 410, BIC = 414), aunque dos ítems se eliminaron por resultar problemáticos. La consistencia de las escalas fue adecuada ($\alpha = .68 - .79$). Se constataron diferencias de género y correlaciones con el rendimiento académico similares a las informadas con otras escalas de ansiedad frente a los exámenes. Se recomienda el empleo de la adaptación obtenida con algunos recaudos.

Palabras clave: *ansiedad ante exámenes; adolescentes; adaptación; análisis factorial, Children's Test Anxiety Scale (CTAS)*

Abstract

The Children's Test Anxiety Scale (CTAS; Wren & Benson, 2004) was adapted to Spanish for its use in adolescents (12 - 17 years old) from Argentina. CTAS is a self report with 30 items grouped in 3 factors named: *task off behaviours, thoughts and autonomic reactions*. The items were directly translated into Spanish, and an equivalence analysis in bilingual sample was done. Two alternative solutions to internal structure of the CTAS were obtained through exploratory factor analysis (N = 360 students, of both sexes, public and private schools, aged 11 to 13 years, M = 12.77, SD = .72), one with three original dimensions and another with four factors, distinguishing worry and self - criticism thoughts. Both models fit was assessed through confirmatory factor analysis (N = 501 students, of both sexes, public and private schools, aged 12 to 17 years, M = 13.98 SD = .87), having the four dimension model the most acceptable index ($\chi^2/df = 1.92$, CFI = .90, TLI = .09, RMSEA = 0.043, AIC = 410, BIC = 414), although two items were eliminated as they were considered problematic. Internal consistency of the scales was adequate ($\alpha = .68 - .79$). Gender differences and correlations with academic performance were corroborated in line with other test anxiety measures. It is recommended to use the adapted version with some precaution.

Keywords: *test anxiety; teenagers; adaptation; factor analysis; Children's Test Anxiety Scale (CTAS)*

* **Correspondencia a:** Luis Alberto Furlan. Laboratorio de Evaluación Psicológica y Educativa (LEPE), Facultad de Psicología. Universidad Nacional de Córdoba. Enrique Barros esq. Enfermera Gordillo. Ciudad Universitaria, Córdoba (5000), República Argentina. Tel.: +543516196136. E-mail: luis.alberto.furlan@unc.edu.ar

Cómo citar: Rivera-Flores, N. B., Furlan, L. A., & Gnani, G. A. (2019). Adaptación de la Children's Test Anxiety Scale (CTAS) para su empleo en adolescentes argentinos. *Revista Evaluar*, 19(2), 73-89. Recuperado de <https://revistas.unc.edu.ar/index.php/revaluar>

Nota de autor: Se agradece a la Lic. María José Ferrero, quien colaboró en la recolección de datos y al equipo directivo del The Red Apple School que colaboró para el estudio de equivalencia en muestra bilingüe.

Introducción

La *ansiedad ante exámenes* puede definirse como una tendencia relativamente estable de algunos individuos a responder con elevados niveles de ansiedad ante situaciones en las que sus aptitudes están siendo evaluadas, el aspecto característico es la preocupación por el posible mal desempeño en la tarea y sus consecuencias aversivas para la autoestima, el estatus o la pérdida de algún beneficio esperado (Cassady & Johnson, 2002; Cizek & Burg, 2006; Zeidner, 2014).

En un principio, la ansiedad frente a los exámenes (en adelante AE) se consideró a los fines teóricos y para su medición una respuesta global y unitaria que incluía manifestaciones cognitivas, somáticas y conductuales (Furlan, 2006). En las décadas de los 60 y 70 del siglo XX hubo dos avances fundamentales en su conceptualización. Por un lado, se distinguió la ansiedad como estado y como rasgo estable de la personalidad (Spielberger, 1966); y por el otro, se diferenciaron dos dimensiones básicas de la experiencia ansiosa, una cognitiva (*preocupación*) y otra de tipo afectivo - fisiológico (*emocionalidad*), lo que dio origen al Modelo Dual (Liebert & Morris, 1967).

Posteriormente, surgieron otros modelos para su medición, debido a que los análisis factoriales realizados sobre los autoinformes de AE mostraron una estructura interna más compleja que la postulada inicialmente. A nivel cognitivo, en el Reactions to Test (RTT; Sarason, 1984) se distinguieron la preocupación y los pensamientos irrelevantes; y en el German Test Anxiety Inventory, (TAI-G, Hodapp, 1995) la preocupación, la interferencia y la falta de confianza. A nivel afectivo-fisiológico, se diferenciaron los sentimientos de tensión y las reacciones somáticas (Sarason, 1984).

En consecuencia, puede afirmarse que la AE es un constructo complejo, multidimensional y

dinámico, para el cual se siguen proponiendo nuevas definiciones y procedimientos de medición (Nyroos & Wiklund-Hörnqvist, 2011) y que por lo general, se ha investigado en poblaciones de jóvenes y adultos, en su mayoría estudiantes universitarios. Sin embargo, la prevalencia de la AE en niños y adolescentes parece ir en aumento, posiblemente debido al incremento de las pruebas en las escuelas y a las presiones asociadas a este (Wren & Benson, 2004).

La AE es un constructo relevante para explicar la variabilidad de rendimiento académico, y existe consenso acerca de sus efectos desfavorables sobre el mismo, como lo señalan algunas revisiones ampliamente citadas en este ámbito (Hembree, 1988; Von der Embse & Hasson, 2012; Zeidner, 1998). Los mecanismos por los cuales se produce dicha influencia comprometen a la atención, la memoria operativa y de largo plazo, el procesamiento de la información y la comprensión lectora, entre otros procesos necesarios para lograr un adecuado desempeño en las evaluaciones (Cassady, 2004; Thomas, Cassady, & Heller, 2017).

Por otro lado, los estudiantes que presentan elevada AE y solicitan asistencia psicológica por este motivo, presentan en la mayoría de los casos otros síntomas del espectro ansioso, obsesiones, baja autoestima, estado de ánimo deprimido, perfeccionismo y elevada sensibilidad interpersonal (Furlan, 2013). Esta evidencia permite suponer que la presencia de elevada AE constituye un indicador de probables dificultades en el ajuste psicológico del estudiante en otros ámbitos de su vida (Furlan, Ferrero, & Gallart, 2014).

Un instrumento para evaluar ansiedad ante exámenes en niños o adolescentes debe ser acorde a las características evolutivas de este grupo etario, como así también a las experiencias escolares a las que hayan estado expuestos previamente. En ese sentido, los instrumentos empleados en

estudiantes universitarios resultan parcialmente adecuados.

El autoinforme más utilizado para niños en edad escolar fue el Test Anxiety Scale for Children (TASC; Sarason, Davidson, Lighthall, Waite, & Ruebush, 1960). El TASC se administra oralmente y se conceptualiza como una única dimensión, medida por treinta elementos utilizando un formato dicotómico (*no/sí*) de respuesta, la escala se consideró apropiada para los niños en los grados 1^{er} al 6^{to}. Sin embargo, este instrumento ha sido cuestionado en las últimas cuatro décadas debido a tres limitaciones: la redacción obsoleta o compleja de los ítems, una definición del dominio desfasada y su dimensionalidad (Wren & Benson, 2004).

Para superar estas dificultades se construyó la Children's Test Anxiety Scale (CTAS, Wren & Benson, 2004) y se reunieron evidencias de su validez y fiabilidad para medir la AE en niños de diversos grupos étnicos. La población seleccionada para su validación (N = 261) fue de estudiantes de los Estados Unidos de América, de 3^{ro} a 6^{to} grado, con edades entre los 8 y 12 años, 53% mujeres y según su origen, 44% afroamericanos, 51% blancos, y 5% asiáticos. La CTAS fue sometida a un análisis de validez de constructo y, mediante análisis factorial exploratorio, se identificaron tres dimensiones. La primera, denominada *pensamiento* ($\alpha = .89$) está conformada por ítems que describen diferentes tipos de cogniciones relacionadas con la actividad en curso o con los resultados esperados. La segunda dimensión, *comportamientos fuera de la tarea* ($\alpha = .78$), posee ítems relacionados con conductas que denotan inquietud o desconexión o dificultades respecto de la tarea. La tercera dimensión, llamada *reacciones autonómicas* ($\alpha = .85$) posee ítems relativos a síntomas físicos de tensión emocional y activación del sistema nervioso autónomo. Un análisis confirmatorio aportó evidencias

adicionales favorables para esta solución de tres factores (Wren & Benson, 2004).

La CTAS despertó el interés en los investigadores de países europeos y asiáticos y por eso, se realizaron adaptaciones para su empleo considerando las características de cada cultura. En un estudio reciente, fue traducida al idioma malayo y obtuvo buenos índices de consistencia interna para sus tres subescalas. No se informaron resultados de su estructura interna para esta adaptación (Ali & Talib, 2015).

También se adaptó al idioma turco (Aydin & Bulgan, 2017) y se incluyeron estudios de estructura interna mediante análisis factorial confirmatorio (N = 1100 estudiantes, 529 mujeres y 571 varones, de entre 9 y 12 años) que apoyaron el modelo de tres factores de la versión original, se obtuvieron buenos índices de ajuste ($\chi^2/sd = 3.97$, RMSEA = .05, SRMR = .05, CFI = .97, GFI = .92, AGFI = .90).

En otra adaptación realizada en Suecia y Finlandia (Nyroos, Korhonen, Linnanmäki, & Svens-Liavåg, 2012) con una muestra binacional (N = 172, 103 niñas y 69 niños, de entre 9 y 10 años; se realizó un análisis factorial exploratorio mediante ecuaciones estructurales, que replicó la estructura original de tres factores, aunque logró mejor ajuste para una versión abreviada de 19 ítems que para la original de 30 ($\chi^2/sd = 1.64$, RMSEA = .06, CFI = .9, TLI = .90)

Otro aspecto considerado a los fines de la validación de la CTAS fue el análisis de las diferencias de género. En los diversos estudios realizados se pudo corroborar lo que plantearon Zeidner (1998), Hembree (1988), Bauermeister, Collazo y Spielberger (1983), Kessler, Sonnega, Bromet, Hughes, & Nelson (1995), Furlan, Casady y Pérez (2009) y Heredia, Piemontesi, Furlan y Hodapp (2008) sobre la diferencia en los niveles de ansiedad frente a los exámenes de acuerdo al género. Las mujeres, independientemente

del entorno cultural y la edad, tienden a informar niveles más altos de AE que los hombres (Lowe & Lee, 2008). Similares resultados encontraron Ali y Talib (2015), quienes informaron que las mujeres puntuaron con niveles más altos de AE en las tres dimensiones, es en la dimensión reacciones autonómicas donde se observó mayor diferencia.

Otras evidencias externas para la validación de las escalas de AE se obtienen mediante el análisis de sus relaciones con algunos indicadores de rendimiento académico (en adelante RA). Al respecto, existen reportes coincidentes de correlaciones negativas y moderadas entre estas variables (Cassady, 2004; Eum & Rice, 2011; Furlan et al, 2009; Gutiérrez-Calvo, 1996; Hembree, 1998; Von der Embse & Hasson, 2012; Zeidner, 2014). Sin embargo, cuando se diferencian la contribuciones relativas de cada dimensión de la AE al RA, se observa que algunas dimensiones cognitivas como la *interferencia* o la *falta de confianza*, realizan las contribuciones más significativas a la variabilidad del RA, en tanto que la *preocupación* tiene efectos menos homogéneos, relacionados generalmente con la forma en que esta dimensión se operacionaliza y las reacciones afectivo-fisiológicas o *emocionalidad* no contribuyen de manera relevante (Piemontesi, Heredia, & Furlan, 2009).

Por otro lado, la AE constituye un factor de riesgo para el bienestar y desarrollo personal, social y académico de niños y adolescentes (Servera, Llabrés & Bornás, 1996; La Greca, Siegel, Wallander, & Walker, 1992), debido a que genera alteraciones específicas en los distintos ámbitos del sujeto, por lo que la ansiedad parece afectar de modo directo a la salud mental en determinadas etapas de la vida, especialmente durante la adolescencia (Axelsson & Ejlertsson, 2002; Bagley & Mallick, 2001). En estudios recientes se advierte sobre la importancia de la AE en relación al

bienestar emocional en la infancia y la necesidad de desarrollar acciones que ayuden a prevenirla y mitigarla (Carsley & Heath, 2019; Segool, Carlson, Goforth, von der Embse, & Barterian, 2013; Von der Embse, Barterian, & Segool, 2013).

En ese sentido, y considerando los hallazgos obtenidos en estudiantes universitarios que indican una elevada comorbilidad de la AE y diversos síntomas mentales (Furlan et al, 2014), se podría esperar que la elevada AE en estudiantes secundarios se asocie a la presencia de otros síntomas mentales o relativos al aprendizaje, cuya presencia podría inferirse a partir de la concurrencia a la asistencia psicológica o psicopedagógica.

A pesar de la gran cantidad de literatura internacional, correlatos y tratamientos para la AE, no hay pruebas que evalúen estos aspectos en escolares de nuestro contexto, lo que dificulta conocer el estado de esta problemática a nivel local. En ese sentido, contar con un instrumento adecuado para la población local permitirá realizar con mayor precisión tareas de diagnóstico, detección temprana de estudiantes en situaciones de riesgo de fracaso escolar y con potencial padecimiento subjetivo, estudios de prevalencia, de relaciones entre este constructo y otros que teóricamente se esperaría que estén vinculados, en población adolescente y en intervenciones cuya eficacia se pretenda evaluar.

Para cubrir esta carencia se realizaron cuatro estudios destinados a adaptar la CTAS para su empleo en adolescentes de Argentina. Cabe aclarar que se eligió la población adolescente, que es unos años mayor que la considerada por los autores que construyeron la CTAS, por diversos motivos. El primero es que los exámenes son un dispositivo de evaluación de uso poco frecuente en las escuelas primarias de la Argentina y es en la secundaria, a la cual se accede a partir de los 12 años, cuando comienzan a ser más habituales. Se consideró necesario que la población para la

cual se adapte la prueba haya tenido experiencias repetidas de exposición ante evaluaciones escritas y orales, ya que la ansiedad frente a los exámenes se define como un rasgo situacional y específico que requiere una tendencia estable de respuesta y surge de ponderar las reacciones emocionales ante una serie de situaciones similares. El segundo motivo para la elección de adolescentes fue la familiaridad con los cuestionarios y las posibilidades de comprender adecuadamente sus consignas y formatos de respuesta. Como se trata de una prueba de autoinforme, es preciso que los estudiantes respondan en forma autónoma y eso requiere una adecuada lectocomprensión, la cual está más garantizada en estudiantes de nivel secundario, en comparación con los de primaria.

El objetivo general que guió los estudios realizados fue elaborar una versión adaptada de la CTAS para su empleo en adolescentes de Argentina. Se establecieron además cuatro objetivos específicos: 1) generar una versión traducida al español de la CTAS y analizar la equivalencia entre esta y su versión original.; 2) reunir evidencias de validez para su estructura interna y evaluar el ajuste de modelos alternativos de medición para la CTAS; y 3) recabar evidencias de validez en base a grupos contrastados, test criterio y de confiabilidad.

Se realizaron estudios para cada objetivo y se formularon las correspondientes hipótesis. 1) Las puntuaciones de las versiones original y traducida se correlacionan en forma positiva y elevada, y las diferencias entre sus medias no son significativas. 2) Los modelos de 3 y 4 factores presentan índices de ajuste adecuados, siendo mejores para el de 4 factores. 3a) Los puntajes de AE son mayores en mujeres que en varones. 3b) Los puntajes de AE son mayores en estudiantes con tratamiento psicológico o psicopedagógico, que en los estudiantes sin tratamiento. 3c) La AE y el rendimiento se correlacionan negativa y

moderadamente.

Metodología

Diseño

Se realizó un estudio de tipo instrumental (Montero & León, 2002), con cuatro etapas relacionadas con: a) la traducción de los ítems, b) el análisis de la estructura interna y confiabilidad de las subescalas, c) el ajuste de los modelos alternativos de medición y, d) las evidencias test criterio y de grupos contrastados.

Participantes

En la primera fase, de traducción y análisis de equivalencia, participaron 12 estudiantes bilingües, pertenecientes a una academia privada de inglés, con edades entre los 13 y los 19 años ($M = 16.17$; $DE = 1.64$). El método de selección fue no probabilístico accidental, cumpliendo con el único requisito de tener idoneidad del idioma inglés para la comprensión del cuestionario.

Para el segundo estudio se empleó una muestra no probabilística accidental, con 360 estudiantes (198 mujeres, 162 varones) de primero y segundo año, con edades entre los 11 y los 13 años ($M = 12.77$; $DE = .72$) de escuelas públicas y privadas de Córdoba.

En el tercer y cuarto estudio participaron 501 alumnos (214 mujeres y 287 varones) de 12 a 17 años ($M = 13.98$ $DE = .87$), seleccionados por muestreo no probabilístico accidental de cuatro colegios públicos y privados de la Provincia de Córdoba.

Instrumentos

Children's Test Anxiety Scale (CTAS; Wren & Benson, 2004). Posee 30 afirmaciones reunidas en tres escalas: pensamientos (13 ítems, $\alpha = .89$), respuestas autonómicas (9 ítems, $\alpha = .85$) y comportamientos fuera de la tarea (8 ítems, $\alpha = .78$), con cuatro opciones de respuesta: *Casi nunca*, *Algunas veces*, *Mayor parte* y *Casi siempre*, que indican la frecuencia de cada afirmación.

Datos socio académicos. Se elaboró un cuestionario ad hoc, con preguntas relativas al rendimiento obtenido en Lengua y Matemáticas durante el primer y segundo trimestre (promedio autoinformado de calificaciones obtenidas en las evaluaciones de cada materia para cada trimestre). También se indagó sobre la concurrencia a tratamiento psicológico o psicopedagógico durante el año en curso (mediante preguntas específicas con respuestas dicotómicas).

Procedimiento

En primera instancia se realizó una traducción directa de los 30 ítems del inglés al español. Posteriormente, se verificó su equivalencia administrando la versión original y la traducida a estudiantes bilingües. La administración se realizó en horarios de clases, con previa solicitud de consentimiento a los padres y con una semana de intervalo entre la versión en inglés y en español. Para analizar la equivalencia entre ambos conjuntos de ítems se calculó la prueba *t* para muestras emparejadas, para el total de los ítems y sus tres subescalas.

Para el primer estudio de estructura interna ($N = 360$), luego de un análisis preliminar orientado a la verificación de los supuestos correspondientes, se realizaron diversos análisis factoriales ex-

ploratorios, utilizando el método de componentes principales y rotación promax. La determinación del número de factores se realizó mediante el test de la pendiente (Cattell, 1966) y el análisis paralelo (Horn, 1965). Se utilizó el programa SPSS (IBM Corporation, 2016) versión 24.

Para un nuevo estudio de estructura interna se realizaron análisis factoriales confirmatorios (AFC) basados en las matrices de covarianza, y las soluciones factoriales fueron generadas con la estimación de máxima verosimilitud. Se utilizaron varios índices para evaluar el ajuste entre los modelos y los datos: chi cuadrado sobre grados de libertad (χ^2/df), índice de ajuste comparativo (CFI), índice de Tucker-Lewis (TLI) y raíz cuadrada media del error de la aproximación (RMSEA). Los siguientes criterios se utilizaron para evaluar el ajuste del modelo: $\chi^2/df \leq 2.0$ (Hair, Anderson, Tatham, & Black, 1995); CFI $\geq .90$; GFI $\geq .90$; y RMSEA $\leq .08$ (Browne & Cudeck, 1993). Para la comparación de múltiples modelos, se utilizaron también el criterio de información de Akaike (AIC) y el criterio de información bayesiano (BIC; mientras más bajos son los valores, mejor es el ajuste). Además, se interpretaron la significación y magnitud de los coeficientes estandarizados de regresión así como la varianza explicada de cada ítem atribuida a su factor latente. Se utilizó el programa Mplus versión 7 (Muthén & Muthén, 2015) para llevar adelante los análisis estadísticos de acuerdo a los objetivos propuestos.

Para los estudios de evidencias de validez en base a grupos contrastados se empleó la prueba *t* para muestras no relacionadas y se estimó el tamaño del efecto (*d* de Cohen), y para analizar las relaciones con el rendimiento (evidencias test-criterio), se calculó el coeficiente de correlación de Pearson.

Aspectos éticos

De acuerdo con los principios éticos vigentes, se informó el objetivo del estudio a los participantes. Al ser menores de edad y debido a que los cuestionarios se administraron en instituciones educativas se solicitó la autorización del personal directivo. Se aseguró la confidencialidad de la información recabada y que los cuestionarios anónimos, así como también el carácter voluntario de su participación. Se comunicó que cualquier sujeto podría abandonar la investigación en el momento que lo deseara. Dado el carácter y tipo de información solicitada a los participantes, no hubo posibilidad de daño en relación con las características de la evaluación. Los instrumentos de evaluación fueron administrados por profesionales idóneos y habilitados para tal fin. Para este estudio se procedió según lo exigen las reglas éticas en vigencia, considerando las pautas propuestas por la [American Psychological Association \(2017\)](#). Para ambos estudios se administró el cuestionario en el horario regular de clases y con autorización previa de los directivos y los padres. La participación fue voluntaria, anónima y los resultados confidenciales.

Resultados

Equivalencia de las versiones original y traducida

Se estimaron la prueba *t* para muestras relacionadas y la correlación de Pearson entre los puntajes totales de las versiones original y traducida de la CTAS. Las medias de los puntajes totales de la versión en inglés y español no difirieron entre sí de forma significativa ($t = -1.51, p = .16$) y presentaron una correlación elevada ($r = .76, p < .001$). Resultados similares se obtuvieron

respecto de las subescalas de pensamientos ($t = -.64, p = .53, r = .87, p < .001$), conductas fuera de la tarea ($t = -1.67, p = .12, r = .36, p = .43$) y reacciones somáticas ($t = -1.62, p = .13, r = .76, p < .001$). Estos resultados indicaron un nivel de equivalencia aceptable entre ambas versiones.

Análisis factorial exploratorio

En primera instancia se realizó un análisis para identificar casos atípicos, siendo $z = \pm 3$ considerados como casos marginales ([Tabachnick & Fidell, 2001](#)), y no se encontraron puntuaciones atípicas. Posteriormente se procedió a verificar el cumplimiento de los supuestos de normalidad; mediante los índices de asimetría y curtosis; y el de multicolinealidad a través de la matriz de correlaciones, resultando ambos satisfactorios.

Los valores de la medida de adecuación muestral Kaiser-Meyer-Olkin ($KMO = 0.88$) y la prueba de esfericidad de Bartlett ($2822.53, gl = 435, sig = p < .001$) indicaron la posibilidad de realizar el AFE, el cual se efectuó a través del método de componentes principales y rotación promax. La regla de Kaiser-Guttman indicó la presencia de seis factores, la interpretación del *scree test* ([Cattell, 1966](#)) y un análisis paralelo ([Horn, 1965](#)) sugirieron la presencia de cuatro. El cuarto factor surgió de la división de la dimensión cognitiva, que puede diferenciarse en *preocupación* (más orientada al resultado y sus implicaciones) y *pensamientos autocríticos* (relacionada con la ejecución). Como el instrumento original estaba conformado por tres factores, se evaluaron dos soluciones alternativas con tres y cuatro factores.

La solución con tres factores explicó el 36.37% y la de cuatro factores el 41.64% total de la varianza, y las escalas obtenidas en ambos casos presentaron valores aceptables de consisten-

cia interna (ver Tabla 1). Sin embargo, surgieron dificultades para interpretar ambas soluciones por las saturaciones de algunos ítems en factores diferentes a los previstos teóricamente. En la de tres factores, los ítems 5, 6 y 19 que describen cogniciones y el ítem 25 de reacciones autonómicas, saturaron en el factor de conductas no relacionadas a la tarea. El ítem 22 presentó una correlación $< .35$ con el factor correspondiente. En cambio, en la solución de cuatro factores el ítem 22, que es una conducta, correlacionó en forma pequeña (.25) con el factor pensamientos autocríticos y el ítem 25 presentó una correlación de .25 con el factor correspondiente, siendo $< .35$ el punto de corte establecido.

Contraste de modelos alternativos de medición mediante análisis factorial confirmatorio (AFC)

El estudio se realizó en una nueva muestra (N = 501) de características similares a la empleada en el análisis exploratorio. Se realizó una modificación en la redacción del ítem número 25 *Tengo que ir al baño* reemplazado por *Siento deseos de ir al baño* para denominar más claramente una reacción fisiológica y no una conducta.

Los modelos propuestos fueron el de tres factores correlacionados para los 30 ítems de acuerdo a la versión original de la CTAS y el de cuatro factores, con los 30 ítems agrupados según lo encontrado en el análisis factorial exploratorio.

Inicialmente se efectuó una exploración inicial de todos los ítems para evaluar valores perdidos, casos atípicos univariados y multivariados, distribuciones normales univariadas y multivariadas y multicolinealidad. No se encontraron valores perdidos. Luego, los casos atípicos univariados y multivariados fueron identificados calculando los puntajes z estándar para cada variable (puntajes $z > 3.29$ fueron considerados atípicos)

y la medida de la distancia de Mahalanobis ($p < .001$). Los resultados de estos análisis fueron satisfactorios, y se detectaron solo tres casos atípicos multivariados, que no fueron descartados. Los valores de asimetría y curtosis fueron de entre -2 y +2 en todos los ítems, menos uno que superó levemente ese límite en la curtosis. Los valores son considerados aceptables para probar una distribución normal univariada (George & Mallery, 2010). Al estimar la multicolinealidad de los ítems mediante correlaciones bivariadas de Pearson se obtuvieron valores aceptables para todos los ítems ($r < .90$).

Estructura interna y consistencia interna del instrumento

La estimación de los modelos presentó resultados de ajuste por debajo de los valores óptimos. Para el modelo de tres factores los resultados fueron claramente insatisfactorios ($\chi^2/df = 2.59$, CFI = .82, TLI = .81, RMSEA = .56, AIC = 413, BIC = 417). Sin embargo, para el modelo de cuatro factores, el desempeño relativo fue algo mejor con algunos valores adecuados y otros cercanos al límite inferior recomendable ($\chi^2/df = 2.16$, CFI = .87, TLI = .86, RMSEA = .048, AIC = 411, BIC = 415). Posteriormente, se exploraron los índices de modificación y para ambos modelos se especificaron tres relaciones entre pares de errores de medición de ítems que guardaban relación empírica debido a la similitud del contenido. Una nueva estimación de los modelos brindó resultados con evidencias favorables para el de cuatro factores. Específicamente, cada uno de los estadísticos cumplió los criterios estipulados para probar un buen ajuste del modelo (Tabla 3).

Las correlaciones entre los factores fueron medianas tanto en el modelo de 3F (.50 - .68) como en el de 4F (.48 - .73). Los coeficientes de

Tabla 1

Análisis factorial exploratorio. Correlaciones ítems-factor de la CTAS para los modelos de 3 y 4 factores.

Ítem	Modelo 3 Factores			Modelo 4 Factores			
	P	RA	CFT	RA	PP	PA	CFT
Me pregunto si aprobaré (1)	.612				.647		
Mi corazón late rápido (2)		.700		.714			
Miro alrededor del aula (3)			.681				.637
Me siento nervioso (4)		.576		.589			
Creo que voy a obtener una mala calificación (5)			.385			.826	
Es difícil para mí recordar las respuestas (6)			.525			.609	
Juego con mi lapicera (7)			.519				.536
Siento caliente la cara (8)		.631		.641			
Me preocupa fallar (9)	.631				.592		
Mi estómago se siente raro (10)		.537		.555			
Me preocupa estar haciendo algo mal (11)	.613				.570		
Verifico el tiempo (12)			.406				.537
Pienso cual será mi nota (13)	0.647				.699		
Me parece difícil quedarme quieto (14)			.439				.446
Me pregunto si mis respuestas son correctas (15)	.552				.678		
Creo que debería haber estudiado más (16)	.396					.778	
Me duele la cabeza (17)		.596		.602			
Miro a las otras personas (18)			.733				.658
Creo que la mayoría de mis respuestas son incorrectas (19)			.486			.806	
Me hace calor (20)		.634		.644			
Me preocupa lo difícil que es la prueba (21)	.484				.393		
Trato de terminar rápido (22)			.323			.317	
Me tiembla la mano (23)		.652		.650			
Pienso en lo que sucederá si fracaso (24)	.468				.432		
Tengo que ir al baño (25)			.380				.276
Muevo los pies (26)			.474				.548
Pienso en lo mal que lo estoy haciendo (27)	.502				.441		
Me siento asustado (28)		.570		.574			
Me preocupa lo que dirán mis padres (29)	.559				.476		
Me quedo mirando fijamente (30)			.446				.358
A	.789	.795	.766	.795	.758	.750	.700
F1							
F2	X						
F3	X	X					
F4	X	X	X				

regresión estandarizados entre los factores y los ítems (beta) fueron superiores a .35, a excepción de los ítems 12 (*Verifico el tiempo*; .27 en 3F y .25 en 4F) y 25 (*Tengo que [o siento ganas de] ir al baño*; .34 en 3F y 4F). Por otra parte, la varianza total explicada por los modelos de medición para cada uno de los ítems (r cuadrado) fue mayor a .10, a excepción del ítem 12 (.07 en 3F y .06 en

4F).

La consistencia interna de las escalas en el Modelo de 3F alcanzó los siguientes valores: *pensamientos* (13 ítems, $\alpha = .85$), *reacciones autonómicas* (9 ítems, $\alpha = .78$) y *conductas fuera de la tarea* (8 ítems, $\alpha = .67$ y si se elimina el ítem 12 mejora a $\alpha = .68$). En el Modelo de 4F:

Tabla 2 Análisis Factorial Confirmatorio. Coeficientes de regresión estandarizados y r^2 de los ítems de la CTAS para los modelos de 3 y 4 factores.

Ítem	Modelo 3 Factores				Modelo 4 Factores				
	P	RA	CFT	r^2	PP	RA	CFT	PA	r^2
Me pregunto si aprobaré (1)	.381			.145	.423				.179
Mi corazón late rápido (2)		.565		.319		.571			.326
Miro alrededor del aula (3)			.399	.159			.409		.168
Me siento nervioso (4)		.575		.331		.594			.353
Creo que voy a obtener una mala calificación (5)	.399			.301				.681	.463
Es difícil para mí recordar las respuestas (6)	.525			.503				.581	.337
Juego con mi lapicera (7)			.450	.276			.463		.214
Siento caliente la cara (8)		.490		.240		.449			.202
Me preocupa fallar (9)	.612			.374	.666				.374
Mi estómago se siente raro (10)		.577		.333		.580			.337
Me preocupa estar haciendo algo mal (11)	.664			.441	.696				.485
Verifico el tiempo (12)			.268	.072			.252		.064
Pienso cual será mi nota (13)	.407			.166	.430				.185
Me parece difícil quedarme quieto (14)			.609	.370			.606		.367

Me pregunto si mis respuestas son correctas (15)	.510		.260		.696			.298
Creo que debería haber estudiado más (16)	.545		.297			.623		.388
Me duele la cabeza (17)		.508	.202			.498		.248
Miro a las otras personas (18)			.468	.219			.483	.233
Creo que la mayoría de mis respuestas son incorrectas (19)	.612			.374			.771	.594
Me hace calor (20)		.570		.325		.540		.292
Me preocupa lo difícil que es la prueba (21)	.628			.395		.613		.395
Trato de terminar rápido (22)			.356	.126			.361	.130
Me tiembla la mano (23)		.534		.285		.530		.281
Pienso en lo que sucederá si fracaso (24)	.591			.349		.593		.352
Tengo que ir al baño (25)		.339		.115		.336		.113
Muevo los pies (26)			.517	.267			.504	.254
Pienso en lo mal que lo estoy haciendo (27)	.625			.390			.651	.424
Me siento asustado (28)		.631		.398		.639		.408
Me preocupa lo que dirán mis padres (29)	.482			.232		.493		.243
Me quedo mirando fijamente (30)			.503	.253			.499	.249
F1	X	.702	.613		X	.745	.531	.531
F2	X	X	.540		X	X	.530	.517
F3	X	X	X		X	X	X	.582
F4	X	X	X		X	X	X	X

pensamientos de preocupación (8 ítems $\alpha = .79$), *pensamientos autocríticos* (5 ítems $\alpha = .79$), *reacciones autonómicas* (9 ítems, $\alpha = .78$) y *conductas fuera de la tarea* (8 ítems, $\alpha = .67$ y si se elimina el ítem 12 mejora a $\alpha = .68$). En función de los resultados obtenidos se considera adecuado eliminar el ítem 12 y mantener el instrumento con 29 ítems distribuidos en 4 escalas.

Evidencias externas de validez, diferencias de género y test criterio

La prueba *t* para muestras independientes indicó la existencia de diferencias significativas en las medias de *reacciones autonómicas* ($p < .001$), *preocupación* ($p < .01$) y AE total ($p < .01$), que

eran mayores en mujeres. Por otro lado, no se observaron diferencias significativas en *conductas fuera de la tarea* y *pensamientos autocríticos*.

Respecto de los tratamientos psicológicos, no se constataron diferencias significativas en los niveles de AE en general, ni en ninguna de sus dimensiones, entre quienes concurren a los mismos ($n = 25$) y quienes no lo hacen ($n = 460$). Los resultados para tratamientos psicopedagógicos fueron similares ($n = 18$ que asisten).

En cuanto a las evidencias test criterio en relación al rendimiento académico, se obtuvieron correlaciones negativas y moderadas ($r = .30 - .33$) entre la escala de *pensamientos autocríticos* y las calificaciones parciales en Matemáticas y Lengua. La AE total correlacionó en forma pequeña y negativa ($r = .17 - .22$) con ambos indicadores de rendimiento académico. A diferencia de los pens-

Tabla 3

Ajuste de los 2 modelos de medición de Ansiedad frente a los Exámenes re especificados.

Modelos	χ^2	df	χ^2/df	CFI	TLI	RMSEA	AIC	BIC
Tres factores	892.4***	399	2.23	.86	.85	.050	411	415
Cuatro factores	761.1***	369	1.92	.90	.89	.043	410	414

amientos autocríticos, las otras tres subescalas de AE no presentaron correlaciones significativas.

Discusión

El proceso de adaptación de la CTAS para su utilización en adolescentes de Argentina permitió obtener una medida de autoinforme que, en términos generales, presenta propiedades satisfactorias de validez y confiabilidad. Respecto de su versión original en inglés, la versión adaptada al castellano posee elevada similitud y a juzgar por los resultados obtenidos en los análisis las correlaciones (elevadas) y diferencias entre grupos (no significativas) su grado de equivalencia

puede considerarse aceptable (Hipótesis 1).

En los estudios sobre su estructura interna, la representación el constructo evaluado logra un mejor ajuste con cuatro dimensiones y no con tres como era la propuesta original (Hipótesis 2). Esta subdivisión del componente cognitivo, diferenciando pensamientos de preocupación y pensamientos autocríticos, resulta potencialmente útil a los fines de investigación y aplicados, y coincide con lo propuesto por diversos instrumentos que miden AE en adultos, como el RTT (Sarason, 1984) que distingue preocupación y pensamientos irrelevantes o el TAI-G (Hodapp, 1995) que diferencia preocupación de interferencia.

Estudios realizados localmente con una

Tabla 4

Evidencias de validez en base a grupos contrastados por género.

	Mujeres n = 241	Varones n = 272	t (Sig.)	d
AE total	2.32 (.54)	2.16 (.47)	3.41 **	.31
Preocupación	2.36 (.59)	2.19 (.57)	3.22 **	.29
Autocrítica	2.37 (.86)	2.24 (.70)	.36 n.s	.16
Reacciones Autonómicas	2.05 (.66)	1.74 (.59)	5.45 **	.49
Conductas fuera de la tarea	2.46 (.69)	2.35 (.52)	1.73 n.s	.18

Nota. *** $p < .001$. ** $p < .01$ * $p < .05$

versión adaptada del TAI-G (GTAI-AR; Piemontesi, Heredia, & Furlan, 2012) demostraron que cada dimensión cognitiva del constructo AE se vincula de manera diferente con otras variables, como el rendimiento, el afrontamiento o la autoeficacia para el aprendizaje. La preocupación se vincula de manera positiva con el afrontamiento de aproximación o de orientación al problema, motivando acciones que disminuyan la probabilidad de fracaso, en tanto que la interferencia se asocia con mayor rumiación, autoculpa y menor rendimiento (Piemontesi, Heredia, Furlan, Sánchez-Rosas, & Martínez, 2012).

En ese sentido, la diferenciación de dos tipos de cogniciones que surgió a partir de los análisis factoriales de la versión adaptada constituye un aporte adicional de valor para comprender las relaciones que cada modalidad de pensamiento establece con otros constructos. Desde el punto de vista temporal, la *preocupación* estaría evaluando pensamientos anticipatorios que pueden motivar el uso de estrategias auxiliares que prevengan el fracaso (Gutiérrez-Calvo, 1996) mientras que los *pensamientos autocríticos*, de carácter más retrospectivo, ejercerían interferencia en la realización de las tareas y a nivel motivacional tendrían un

efecto menos favorable, lo que provocaría desánimo.

Con respecto a las evidencias externas de validez (Hipótesis 3a) los resultados corroboraron parcialmente las diferencias esperadas en los niveles de AE de acuerdo al sexo, que en el caso de la emocionalidad y la preocupación aparecen más acentuadas en las mujeres (Ali & Talib, 2015; Lowe & Lee, 2008). Sin embargo, no se observaron diferencias en la escala de *conductas fuera de la tarea* y *pensamientos autocríticos*. Estas discrepancias entre los resultados observados según la dimensión considerada sugieren la necesidad de preservar la conceptualización y medición de la AE como constructo multidimensional, ya que cada uno de sus componentes presenta relaciones específicas con otras variables.

En relación a la sensibilidad de la AE para inferir la presencia de otras condiciones que requieran tratamiento psicológico o psicopedagógico, las diferencias entre grupos no fueron significativas y, en consecuencia, se debe rechazar la hipótesis 3b. Probablemente, el reducido tamaño de la muestra que recibía tratamiento (no superó el 10% del total) sea una limitación para el análisis propuesto y el indicador elegido no sea repre-

sentativo del fenómeno que se pretendía estimar en forma indirecta.

Finalmente, y tal cual se postuló (Hipótesis 3c) la AE presentó una correlación negativa y moderada con el RA al ser considerada como constructo general (Cassady, 2004; Eum & Rice, 2011; Furlan et al., 2009; Hembree, 1998; Von der Embse & Hasson, 2012; Zeidner, 2014). Sin embargo, solo algunas de sus dimensiones presentaron una correlación significativa con el rendimiento y otras no estarían relacionadas con el mismo, como se ha demostrado en estudios realizados con otras medidas multidimensionales de AE (Piemontesi et al., 2012).

Limitaciones

Los estudios de estructura interna de la versión adaptada de la CTAS se llevaron a cabo en muestras de alumnos con una media de edad superior a la considerada en los estudios originales. Este es un factor que podría suponer limitaciones al momento de comparar los resultados de los análisis realizados en las diferentes versiones de la escala. Los motivos para esta decisión fueron expuestos previamente. Se puede agregar que hay menores probabilidades de sesgos en las respuestas derivadas de problemas en la comprensión de las consignas cuando un instrumento construido para niños se administra con adolescentes que si el caso fuera a la inversa. Se espera, de hecho, que la comprensión lectora mejore con la edad y la mayor escolarización y por eso no se considera que esta modificación de la población meta haya podido inducir efectos contraproducentes.

Los niveles de consistencia interna de las escalas obtenidas han sido aceptables mas no plenamente satisfactorios. Mejorarlos supondría

reemplazar algunos ítems por otros nuevos que debieran construirse en base a una revisión minuciosa de la definición operacional de cada dimensión. Por otro lado, esta labor excede lo propuesto inicialmente y daría lugar a un instrumento diferente al original, que no permitiría su uso en estudios transculturales.

Otra limitación importante radica en las variables elegidas para reunir evidencias de validez. Se han considerado las diferencias de género e indicadores autoinformados de rendimiento, lo que arrojó resultados similares a los informados con otras medidas de AE.

Sin embargo, no se han evaluado otras variables teóricamente vinculadas con la AE, como la ansiedad general estado - rasgo, ansiedad social, la dimensión neuroticismo o estabilidad emocional de la personalidad, la autoestima o el perfeccionismo. La presencia de sintomatología mental, que frecuentemente se halla en comorbilidad con la AE (Furlan et al., 2014), no se evaluó en forma directa, sino que se estimó en forma indirecta a partir de la concurrencia a servicios de psicología o psicopedagogía, sin que se pudieran constatar diferencias significativas en la AE entre quienes concurren a los mismos y quienes no lo hacen.

Estudios posteriores podrían considerar la inclusión de algunas de estas variables, como así también otras relativas a los contextos de socialización, como pueden ser las exigencias parentales, el estilo de instrucción del docente o los climas áulicos.

Conclusiones

La adaptación al español de la CTAS aporta un nuevo instrumento en condiciones para su utilización en las escuelas secundarias de gestión estatal o privada del ámbito local. Su empleo en otras regiones de la Argentina deberá hacerse con

precaución, especialmente en lo relativo a aspectos característicos del lenguaje de uso frecuente en los adolescentes, de modo de verificar la comprensión de los ítems.

Esta herramienta de evaluación tiene utilidad aplicada con fines de diagnóstico, de tipo clínico individual, como así también para explorar el clima emocional en grupos de estudiantes. Detectar tempranamente los elevados niveles de AE en los grupos de adolescentes permite implementar estrategias psicoeducativas tendientes a disminuirla, mediante el entrenamiento en estrategias de aprendizaje auto regulado y preparación para rendir, estrategias de regulación emocional y también mediante intervenciones con padres y docentes para modificar factores contextuales asociados con su incremento.

Referencias

- Ali, N. S. M., & Talib, R. (2015). Test anxiety in school settings: Implication on teachers. *Indonesian Journal of Educational Review*, 2(2), 67-75. Recuperado de <http://pps.unj.ac.id/journal/ijer/index>
- American Psychological Association. (2017). *Ethical Principles of Psychologists and Code of Conduct*. Recuperado de <http://www.apa.org/ethics/code/index.aspx>
- Axelsson, L., & Ejlertsson, G. (2002). Self-reported health, self-esteem and social support among young unemployed people: A population-based study. *International Journal of Social Welfare*, 11(2), 111-119. doi: 10.1111/1468-2397.00205
- Aydın, U., & Bulgan, G. (2017). Çocuklarda Sınav Kaygısı Ölçeği'nin Türkçe Uyarlaması. *İlköğretim Online*, 16(2), 887-899. doi: 10.17051/ilkonline.2017.304742
- Bagley, C., & Mallick, K. (2001). Normative data and mental health construct validity for the Rosenberg Self-Esteem Scale in British adolescents. *International Journal of Adolescence and Youth*, 9(2-3), 117-126. doi: 10.1080/02673843.2001.9747871
- Bauermeister, J., Collazo, J., & Spielberger, C. (1983). The construction and validation of the Spanish form of the Test Anxiety Inventory: Inventario de Auto Evaluación sobre Exámenes (IDASE). En C. D. Spielberger & R. Diaz Guerrero (Eds.). *Cross-cultural Anxiety* (pp. 67-85). Washington: McGraw-Hill
- Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. En: K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Beverly Hills, CA: Sage.
- Carsley, D., & Heath, N. L. (2019). Evaluating the effectiveness of a mindfulness coloring activity for test anxiety in children. *The Journal of Educational Research*, 112(2), 143-151. doi: 10.1080/00220671.2018.1448749
- Cassady, J. C. (2004). The impact of cognitive test anxiety on text comprehension and recall in the absence of external evaluative pressure. *Applied Cognitive Psychology*, 18(3), 311-325. doi: 10.1002/acp.968
- Cassady, J. C., & Johnson, R. E. (2002). Cognitive test anxiety and academic performance. *Contemporary Educational Psychology*, 27(2), 270-295. doi: 10.1006/ceps.2001.1094
- Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, 1(2), 245-276. doi: 10.1207/s15327906mbr0102_10
- Cizek, G. J., & Burg, S. S. (2006). *Addressing test anxiety in a high-stakes environment: Strategies for classrooms and schools*. Thousand Oaks, CA: Corwin.
- Eum, K., & Rice, K. G. (2011). Test anxiety, perfectionism, goal orientation, and academic performance. *Anxiety, Stress, & Coping*, 24(2), 167-178. doi: 10.1080/10615806.2010.488723
- Furlan, L. (2006). Ansiedad ante los exámenes. Qué se evalúa y cómo? *Evaluar*, 6, 32-51. Recuperado de <http://revistas.unc.edu.ar/index.php/revaluar>
- Furlan, L. (2013). Eficacia de una intervención para disminuir la ansiedad frente a los exámenes en estudiantes universitarios argentinos. *Revista Colombiana de Psicología*, 2(1), 75-89. doi: 10.15446/rcp

- Furlan, L. A., Cassady, J. C., & Pérez, E. R. (2009). Adapting the Cognitive Test Anxiety Scale for use with Argentinean university students. *International Journal of Testing*, 9(1), 3-19. doi: [10.1080/15305050902733448](https://doi.org/10.1080/15305050902733448)
- Furlan, L. A., Ferrero, M. J., & Gallart, G. (2014). Ansiedad frente a los exámenes, procrastinación y síntomas mentales en estudiantes universitarios. *Revista Argentina de Ciencias del Comportamiento*, 6(3), 31-39. Recuperado de <https://revistas.unc.edu.ar/index.php/racc>
- George, D., & Mallery, P. (2010). *SPSS for Windows step by step: A simple guide and reference, 17.0 Update* (10ª ed.). Boston, MA: Pearson.
- Gutiérrez-Calvo, M. (1996). Ansiedad y deterioro cognitivo: Incidencia en el rendimiento académico. *Ansiedad y Estrés*, 2(2-3), 173-194. Recuperado de <https://dialnet.unirioja.es>
- Hair, J. F. J., Anderson, R. E., Tatham, R. L., & Black, W. C. (1995). *Multivariate data analysis* (4ª ed.). Saddle River, NJ: Prentice Hall.
- Hembree, R. (1988). Correlates, causes, effects, and treatment of test anxiety. *Review of Educational Research*, 58(1), 47-77. doi: [10.2307/1170348](https://doi.org/10.2307/1170348)
- Heredia, D., Piemontesi, S., Furlan, L., & Hodapp, V. (2008). Adaptación del Inventario Alemán de Ansiedad frente a los Exámenes: GTAI-A. *Revista Evaluar*, 8(1), 46-60. Recuperado de <https://revistas.unc.edu.ar/index.php/revaluar>
- Hodapp, V. (1995). The TAI-G: A multidimensional approach to the assessment of test anxiety. En C. Schwarzer & M. Zeidner (Eds.), *Stress, Anxiety, and Coping in Academic Settings*, (pp. 95-130). Francke, Tübingen.
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30(2), 179-185. doi: [10.1007/bf02289447](https://doi.org/10.1007/bf02289447)
- IBM Corporation. (2016). IBM SPSS Statistics for Windows, Version 24.0 [software de cómputo]. Armonk, NY: IBM.
- Kessler, R. C., Sonnega, A., Bromet, E., Hughes, M., & Nelson, C. B. (1995). Posttraumatic stress disorder in the national comorbidity survey. *Archives of General Psychiatry*, 52, 1048-1060.
- La Greca, A. M. E., Siegel, L. J., Wallander, J. L., & Walker, C. (1992). *Stress and coping in child health*. Nueva York, NY: Guilford.
- Liebert, R. M., & Morris, L. W. (1967). Cognitive and emotional components of test anxiety: A distinction and some initial data. *Psychological Reports*, 20(3), 975-978. doi: [10.2466/pr0.1967.20.3.975](https://doi.org/10.2466/pr0.1967.20.3.975)
- Lowe, P. A., & Lee, S. W. (2008). Factor structure of the Test Anxiety Inventory for Children and Adolescents (TAICA) scores across gender among students in elementary and secondary school settings. *Journal of Psychoeducational Assessment*, 26(3), 231-246. doi: [10.1177/0734282907303773](https://doi.org/10.1177/0734282907303773)
- Montero, I., & León, O. G. (2002). Clasificación y descripción de las metodologías de investigación en Psicología. *International Journal of Clinical and Health Psychology*, 2(3), 503-508. Recuperado de <http://www.aepc.es/ijchp/busca.php?coid=Espa%Flol>
- Muthén, L. K., & Muthén, B. O. (1998-2015). *Mplus User's Guide* (7ª ed.). Los Angeles, CA: Muthén & Muthén.
- Nyroos, M., Korhonen, J., Linnanmäki, K., & Svens-Liavåg, C. (2012). A cross-national comparison of test anxiety in Swedish and Finnish grade 3 pupils: Measured by the CTAS. *Education Inquiry*, 3(4), 615-636. doi: [10.3402/edui.v3i4.22057](https://doi.org/10.3402/edui.v3i4.22057)
- Nyroos, M., & Wiklund-Hörnqvist, C. (2011). Introducing national test in Swedish primary education: Implications for test anxiety. *Electronic Journal of Research in Educational Psychology*, 9(25), 995-1022. doi: [10.25115/ejrep.v9i25.1466](https://doi.org/10.25115/ejrep.v9i25.1466)
- Piemontesi, S., Heredia, D., & Furlan, L. (2009). Correlatos de la ansiedad ante los exámenes: Una aproximación a la teoría de la reducción en la eficiencia. *Revista Tesis*, 2, 74-86. Recuperado de <https://revistas.unc.edu.ar/index.php/tesis>
- Piemontesi, S., Heredia, D., & Furlan, L. (2012). Propiedades psicométricas de la versión en español revisada del Inventario Alemán de Ansiedad ante Exámenes

- (GTAI-AR) en estudiantes argentinos. *Universitas Psychologica*, 11(1), 177-186.
- Piemontesi, S., Heredia, D., Furlan, L., Sánchez, J., & Martínez, M. (2012). Ansiedad ante los exámenes y estilos de afrontamiento ante el estrés académico en estudiantes universitarios. *Anales de Psicología*, 28(1), 89-96.
- Sarason, I. G. (1984). Stress, anxiety, and cognitive interference: Reactions to tests. *Journal of Personality and Social Psychology*, 46, 929-938.
- Sarason, S. B., Davidson, K. S., Lighthall, F. F., Waite, R. R., & Ruebush, B. K. (1960). *Anxiety in elementary school children: A report of research*. Hoboken, NJ: John Wiley & Sons. doi: [10.1037/14349-000](https://doi.org/10.1037/14349-000)
- Segool, N. K., Carlson, J. S., Goforth, A. N., von der Embse, N., & Barterian, J. A. (2013). Heightened test anxiety among young children: Elementary school students' anxious responses to high-stakes testing. *Psychology in the Schools*, 50(5), 489-499. doi: [10.1002/pits.21689](https://doi.org/10.1002/pits.21689)
- Servera, M.; Llabrés, J., & Bornás, X. (1996). Prevención de la ansiedad en escolares. *Ansiedad y Estrés*, 2(2-3), 283-295.
- Spielberger, C. D. (1966). Theory and research in anxiety. En C. D. Spielberger (Ed.), *Anxiety and Behavior* (pp. 3-20). Nueva York, NY: Academic. doi: [10.1016/b978-1-4832-3131-0.50006-8](https://doi.org/10.1016/b978-1-4832-3131-0.50006-8)
- Tabachnick, B., & Fidell, L. (2001). *Using multivariate statistics* (4ª ed.). Boston: Allyn and Bacon.
- Thomas, C. L., Cassady, J. C., & Heller, M. L. (2017). The influence of emotional intelligence, cognitive test anxiety, and coping strategies on undergraduate academic performance. *Learning and Individual Differences*, 55, 40-48. doi: [10.1016/j.lindif.2017.03.001](https://doi.org/10.1016/j.lindif.2017.03.001)
- Von der Embse, N., Barterian, J., & Segool, N. (2013). Test anxiety interventions for children and adolescents: A systematic review of treatment studies from 2000-2010. *Psychology in the Schools*, 50(1), 57-71. doi: [10.1002/pits.21660](https://doi.org/10.1002/pits.21660)
- Von der Embse, N., & Hasson, R. (2012). Test anxiety and high-stakes test performance between school settings: Implications for educators. *Preventing school failure: Alternative Education for Children and Youth*, 56(3), 180-187. doi: [10.1080/1045988X.2011.633285](https://doi.org/10.1080/1045988X.2011.633285)
- Wren, D. G., & Benson, J. (2004). Measuring test anxiety in children: Scale development and internal construct validation. *Anxiety, Stress, & Coping*, 17(3), 227-240. doi: [10.1080/10615800412331292606](https://doi.org/10.1080/10615800412331292606)
- Zeidner, M. (1998). *Test anxiety: The state of the art*. Nueva York, NY: Plenum.
- Zeidner, M. (2014). Anxiety in education. En R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International Handbook of Emotions in Education* (pp. 265-288). New York, NY: Routledge. doi: [10.4324/9780203148211.ch14](https://doi.org/10.4324/9780203148211.ch14)