

Validación de contenido de una versión culturalmente adaptada al contexto venezolano de la Escala de Perfeccionismo Infantil de Oros

Leonardo Andrés Aguilar-Durán*, Marilyn Yisneida Castellanos-López**

Resumen

Introducción: a pesar de su importancia, el reporte formal y pormenorizado de la validez de contenido de los instrumentos psicométricos ha sido frecuentemente descuidado. En un trabajo previo, la Escala de Perfeccionismo Infantil de Oros (2003), de origen argentino, fue adaptada lingüística y culturalmente al contexto venezolano. **Objetivo:** el propósito del presente estudio fue estimar la validez relacionada con el contenido de la versión venezolana de la escala antedicha a través del criterio de jueces. **Metodología:** un grupo de expertos en el tema del perfeccionismo juzgó si existe correspondencia entre lo que parece que mide cada reactivo del instrumento y la dimensión a la que pertenece; para cuantificar el acuerdo interjueces se usó el denominado Índice de Validez Factorial (FVI). **Resultados:** la escala total y las llamadas “Autodemandas”, que se refieren a pensamientos perfeccionistas, obtuvieron índices satisfactorios (FVI > 0,80). Hubo menos acuerdo con los ítems de la dimensión “Reacciones ante el fracaso”, en parte atribuible a factores de diferenciación cultural que influyen en la expresión emocional. **Conclusiones:** ahora puede proseguirse con el resto de la validación psicométrica de la prueba. Otros estudios que persigan un objetivo similar al de este trabajo podrían emular el procedimiento aquí expuesto.

Palabras clave: juicio de expertos, niños venezolanos, perfeccionismo infantil, psicometría, validez de contenido.

* Licenciado en Psicología (Opción Psicología Clínica), Universidad Central de Venezuela. Escuela de Psicología, Facultad de Humanidades y Educación, Universidad Central de Venezuela, Venezuela.
Correo electrónico: leonardo.aguilar.d@ucv.ve

** Licenciada en Psicología (Opción Psicología Clínica), Universidad Central de Venezuela. Escuela de Psicología, Facultad de Humanidades y Educación, Universidad Central de Venezuela, Venezuela.
Correo electrónico: castellanos.lmarilyn@hotmail.com

Recibido: 9 de marzo del 2016

Aprobado: 30 de mayo del 2016

Cómo citar este artículo: Aguilar-Durán, L. A. y Castellanos-López M. Y. (2016). Validación de contenido de una versión culturalmente adaptada al contexto venezolano de la Escala de Perfeccionismo Infantil de Oros. *Pensando Psicología*, 12(20), 29-41. doi: <http://dx.doi.org/10.16925/pe.v12i20.1561>



Content Validation of a Version of Oros' Child Perfectionism Scale Culturally Adapted to the Venezuelan Context

Abstract

Introduction: Despite its importance, formal, detailed reporting on the content validity of psychometric instruments has often been neglected. In a previous work, the Child Perfectionism Scale by Oros (2003), from Argentina, was linguistically and culturally adapted to the Venezuelan context. *Objective:* The purpose of this study was to estimate content validity of the Venezuelan version of the above scale through the opinion of judges. *Methodology:* A group of experts in the field of perfectionism judged whether there is correspondence between what each item of the instrument appears to measure and the dimension to which they belong. To quantify the inter-judge agreement, the so-called Factorial Validity Index (FVI) was used. *Results:* The total scale and the so-called "Self-Demands", which refer to perfectionist thoughts, achieved satisfactory indices ($fvi > 0.80$). There was less agreement with items in the dimension "Reactions to Failure," partly attributable to cultural differentiation factors influencing emotional expression. *Conclusions:* Now, the rest of the psychometric validation of the test can be completed. Other studies that pursue an objective similar to that of this work could emulate the procedure presented herein.

Keywords: expert opinion, Venezuelan children, child perfectionism, psychometry, content validity.

Validação do conteúdo de uma versão adaptada do ponto de vista cultural ao contexto venezuelano da Escala de Perfeccionismo Infantil de Oros

Resumo

Introdução: embora sua importância, o relatório formal e pormenorizado da validade de conteúdo dos instrumentos psicométricos tem sido frequentemente descuidado. Em um trabalho prévio, a Escala de Perfeccionismo Infantil de Oros (2003), de origem argentina, foi adaptada linguisticamente e culturalmente ao contexto venezuelano. *Objetivo:* o escopo deste estudo foi calcular a validade relacionada com o conteúdo da versão que mede cada reativo do instrumento e a dimensão à qual pertence; para quantificar o acordo entre juízes foi usado o chamado Índice de Validade Fatorial (FVI). *Resultados:* a escala total e as chamadas "Auto-demandas", que se referem a pensamentos perfeccionistas, obtiveram índices satisfatórios ($FVI > 0,80$). Houve menos acordo com os itens da dimensão "Reações perante o Fracasso", em parte pode ser atribuído a fatores de diferenciação cultural que incide na expressão emocional. *Conclusões:* agora pode avançar com a parte restante da validação psicométrica do teste. Outros estudos procurando um escopo similar ao objetivo deste trabalho poderiam imitar o procedimento aqui apresentado.

Palavras chave: juízo de especialistas, crianças venezuelanas, perfeccionismo infantil, psicometria, validade de conteúdo.

Introducción

La calidad de una prueba psicológica se determina mediante sus indicadores de validez y confiabilidad (Cohen y Swerdlik, 2001). La validez se define como “el grado en que una prueba mide lo que pretende medir” (Hogan, 2003, p. 118). Esta es la característica más importante de un instrumento psicométrico y ha sido considerada como el aspecto central de la psicometría (Hogan, 2003; Martínez, 1996). No obstante, dentro de las categorías que están relacionadas con la validez como concepto unitario (Kaplan y Saccuzzo, 2006; Martínez, 1996; Muñiz, 1992), la validez intrínseca o de contenido ha sido la gran olvidada a la hora de reportar formalmente y en detalle las propiedades psicométricas de un instrumento estandarizado de medición.

El concepto de validez de contenido ha sido objeto de múltiples modificaciones desde sus orígenes (Martínez, 1996; Pedrosa et al., 2013; Sireci, 2009). Sin embargo, estos cambios han estado más bien focalizados en las evidencias que este tipo de validez debe presentar, así como en los diferentes métodos para su estudio, lo cual ha mantenido su definición esencialmente estable a lo largo del tiempo. De esta manera, según Nunnally y Bernstein (1995, p. 113), la validez de contenido depende de “la adecuación con la que se muestrea un dominio de contenido específico”. Para Anastasi y Urbina (1998, p. 115), los procedimientos de validación de contenido “comprenden principalmente el examen sistemático del contenido de la prueba para determinar si cubre una muestra representativa del área de conducta que debe medirse”. Así, el propósito de un estudio de validez de contenido sería el de “establecer el grado en que el conjunto de ítems de un test representa adecuadamente un dominio de conductas de interés específico” (Martínez, 1996, p. 335).

Es importante aclarar que, aunque la validez de contenido se ha considerado especialmente pertinente en la valoración de los tests criteriales, por ejemplo, las pruebas de rendimiento académico y ocupacional (Anastasi y Urbina, 1998; Dembo, 1983; Hogan, 2003; Kaplan y Saccuzzo, 2006; Magnusson, 1975; Nunnally y Bernstein, 1995)—, en la actualidad el acuerdo acerca de la importancia que presenta en el proceso de creación (o adaptación) y validación de cualquier instrumento de medida es unánime (Abad et al., 2011; Kane, 2009; Martínez, 1996).

Oros (2003) desarrolló una Escala de Perfeccionismo Infantil, con el objetivo de medir el perfeccionismo insano en los niños argentinos de entre 8 y 13 años de edad. Esta escala es la primera prueba en español que permite explorar el constructo en el caso infantil, y lo hace a través de la evaluación de ocho autodemandas irracionales y ocho emociones asociadas al fracaso de las autodemandas. La autora estudió el poder discriminativo de los ítems, la validez factorial de la escala y su consistencia interna, con lo cual encontró resultados satisfactorios.

Volviendo sobre la validación de contenido, se tiene que esta descansa, sobre todo, en apelaciones a la razón acerca de la adecuación con que ha sido muestreado un contenido importante e incluido en reactivos de prueba (Nunnally y Bernstein, 1995). Es decir, su análisis se apoya más en la lógica que en hallazgos estadísticos o de naturaleza empírica (Kaplan y Saccuzzo, 2006). Es por eso que, de acuerdo con Nunnally y Bernstein (1995), uno de los aspectos principales para asegurar la validez de contenido tiene que ver con el empleo de métodos “sensibles” de construcción de pruebas. En este sentido, como es lo recomendado (Anastasi y Urbina, 1998; Matesanz, 1997; Nunnally y Bernstein, 1995), el procedimiento de construcción de la escala original argentina se inició con actividades exploratorias fundamentalmente abocadas a la búsqueda de material bibliográfico y a las entrevistas con niños, docentes y psicólogos (Oros, 2003).

En atención a la necesidad descrita por Kane (2006) de un dominio bien definido, la revisión de la literatura evidenció que Hewitt y Flett están entre los investigadores que más han aportado a la temática del perfeccionismo, al desarrollar en la década de los noventa una escala para medir el constructo (Hewitt y Flett, 1991), y proveer evidencia consistente según la cual el perfeccionismo puede ser considerado un fenómeno multidimensional. Inicialmente, se redactaron 31 ítems con arreglo a dos de las dimensiones comprendidas por el esquema teórico de Hewitt y Flett (1991): 1) el perfeccionismo orientado hacia sí mismo, el cual refiere una serie de cogniciones y comportamientos basados en objetivos personales muy difíciles de cumplir, autoevaluaciones minuciosas excesivamente exigentes, un elevado nivel de crítica personal, la búsqueda de la perfección y la evitación del fracaso; y 2) el perfeccionismo prescrito socialmente, el cual describe la necesidad de cumplir con las expectativas dictadas por otros significativos, y tiene que ver con la percepción de las

personas que esperan un comportamiento perfecto, por lo cual se propende a complacer tales expectativas con el deseo de ganar la aceptación y el afecto de aquellas. Al mismo tiempo, interesaba explorar si los pensamientos perfeccionistas podían aparecer en determinadas áreas de funcionamiento, tales como la académica, la social o la deportiva. No fue sino a partir de la organización final de los ítems que fueron seleccionados por tener el desempeño psicométrico de mayor calidad que se generaron las dimensiones finales del instrumento (i.e., Autodemandas y Reacciones ante el fracaso).

A fin de evaluar la claridad verbal de los ítems construidos en el contexto argentino, se recurrió a la opinión de 60 niños escolarizados de ambos sexos y de dos docentes de nivel inicial. Los niños debían responder las siguientes preguntas para cada ítem que incluía el instrumento: (a) ¿se entiende?, ¿es claro?, (b) ¿qué significa?, y (c) ¿se puede decir lo mismo de una manera más fácil? Con las respuestas a estos interrogantes fue posible, además de evaluar la claridad verbal, recoger información acerca de la validez de los reactivos, comprobando cómo aquello que los niños interpretaban de los enunciados era precisamente lo que se pretendía medir. A los docentes se les solicitó que leyeran los enunciados y marcaran aquellos que, desde su experiencia, suponían difíciles de entender para los niños. De igual forma, se les solicitó proponer, en el caso que fuese conveniente, una redacción alternativa. Por último, los psicólogos analizaron cuidadosamente los enunciados con el propósito de observar el grado de adecuación con la teoría.

En general, las advertencias y recomendaciones expresadas por los niños y los profesionales facilitaron y orientaron la redacción de los ítems que conformaron la escala original. Una vez concluida esta etapa, se dio un formato al conjunto de ítems y se optó por el estilo de respuesta más adecuado para cada caso. La elección de la forma de respuesta fue una de las dificultades con las que se enfrentó la autora, ya que según la opinión de los expertos lo aconsejable era expresar las autodemandas en el mismo modo en que los niños las verbalizan. Varias de estas autodemandas se expresan en forma negativa como “no puedo cometer errores” o “no debo equivocarme nunca”. Se ha demostrado (Matesanz, 1997; Moreno et al., 2004) que cuando los ítems se redactan en forma negativa, no es conveniente presentar como alternativa de respuesta el término “no”, ya que podría establecerse una doble negación que confundiría a los niños.

Por esta razón, la escala se dividió en dos apartados con varias alternativas de respuesta: los niños responden con “sí”, “a veces” y “no” a la mayoría de los ítems, excepto aquellos que comienzan con el término “no”, los cuales se responden con “lo pienso mucho”, “lo pienso a veces” y “no lo pienso”. Si bien esta última alternativa también incluía la negación, se notó que redactada de este modo resultaba más comprensible.

La importancia de estudiar el perfeccionismo estriba en que es una variable de personalidad que favorece la aparición de patología psicológica en adultos, así como hay evidencias de que juega un papel importante en las dificultades psicológicas que los niños experimentan (Lozano, Llanos, Lozano y Valor-Segura, 2014; Lozano, Valor-Segura, Llanos y Lozano, 2014). Recientemente, los autores de este artículo se percataron del vacío local de conocimiento respecto al perfeccionismo infantil, ya que no encontraron investigaciones sobre la variable en niños o adolescentes venezolanos. De allí que un primer paso hacia la obtención de una herramienta sensible a la medición del perfeccionismo en la población infantil venezolana lo constituyó la adaptación cultural de la escala de Oros (2003) al referido contexto. En este proceso participaron lingüistas, docentes de educación primaria y niños de la población diana. Justamente, una muestra especial de atención recibió la realización de entrevistas a niños que formaban parte de la población a la que va dirigida la escala, quienes son los que viven el fenómeno y podían hacer “hablar los reactivos”, con lo cual se verificó que estos eran comprensibles y cercanos a la cultura de la población objetivo, de manera que no fue necesario modificar la expresión lingüística de los ítems más allá de las sugerencias que ya habían apuntado los expertos en lengua y los educadores a los que se había consultado.

Además del análisis racional y de la definición operativa del dominio (Sireci, 2003) que hasta ahora se ha presentado, los autores especializados (Anastasi y Urbina, 1998; Hogan, 2003; Kaplan y Saccuzzo, 2006; Muñiz, 1992; Nunnally y Bernstein, 1995) proponen estimar la validez de contenido mediante alguno de los métodos basados en el juicio de expertos, los cuales se caracterizan por contar con un número de jueces que, conociendo las dimensiones que se pretenden evaluar, valoran y asignan cada uno de los ítems a dichas dimensiones (Sireci, 1998a; 1998b). En el presente estudio, los ítems de la versión venezolana de la Escala de Perfeccionismo Infantil fueron revisados por expertos en el constructo examinado con el fin de hacer un análisis de congruencia entre los reactivos y las dimensiones a las que pertenecen. De

esta manera, fue posible proveer indicios de que los ítems representaban lo que se pensaba medir, según la propuesta que había advenido a partir del análisis de la estructura factorial de la escala en el contexto argentino.

Método

Tipo de investigación

Puesto que este estudio propone analizar una de las categorías de la validez de un instrumento de medición, es preciso considerarlo como de carácter instrumental (Carretero-Dios y Pérez, 2007; Montero y León, 2005, 2007).

Participantes

Debido a que la apropiada selección de los expertos supone una cuestión primordial a la hora de establecer este tipo de validez (Martínez, 1996; Pedrosa et al., 2013), resulta fundamental analizar sus características y su experiencia con relación al constructo tratado (véase la tabla 1): fueron convocados vía correo electrónico 68 expertos españoles y latinoamericanos, se obtuvo respuesta de 31 de ellos (46%): 17 de España (55%), cinco de Argentina (16%), cuatro de Venezuela (13%), tres de México (10%), uno de Chile (3%) y uno de Paraguay (3%). Estos expertos internacionales provienen de varias áreas del conocimiento relacionadas, tales como la psicología

(principalmente la psicología clínica y de la salud), la psiquiatría y la psicopatología, y la educación (en la cual se destacan los campos de la educación física y la educación especial); algunos de los clínicos cuentan con formación terapéutica específica (concretamente en terapia conductual y terapia cognitiva, ambos enfoques incluso). Otro dato es que 28 de los expertos tienen estudios de posgrado (90%), 21 cuentan con un doctorado (68%), y tres están en proceso de obtenerlo (10%). Además, cinco (16%) poseen doble titulación de pregrado (i.e., psicología con filosofía, eugenesia integral, terapia ocupacional y educación).

Ahora bien, lo que les daba la condición de expertos a los profesionales contactados era que fueran autores o coautores de, por lo menos, un trabajo académico (e.g., artículo publicado, tesis), teórico o empírico, el cual involucrara al perfeccionismo como variable de estudio fundamental. Un buen número de estos trabajos —según el marco de investigación al que se adscribieron— se realizaron en muestras de niños y adolescentes. Además del dominio de la teoría subyacente al constructo del perfeccionismo, varios de los expertos tenían cierta experiencia en su tratamiento como parte de la práctica clínica habitual o derivada de la misma investigación; por ejemplo, procurando incluir en las discusiones de sus estudios recomendaciones más o menos vinculadas con el abordaje clínico del perfeccionismo en el caso infantojuvenil. Adicionalmente, varios expertos tenían conocimiento en relación con la construcción y la adaptación de escalas de evaluación psicológica, algunas específicas al perfeccionismo.

Tabla 1

Datos de interés sobre los expertos que participaron en la validación de contenido de la versión venezolana de la escala

Núm. Experto	País	Formación académica	Marco de investigación
1	México	Lic. en Psicología, Maestra en Psicología Clínica, Dra. en Psicología de la Salud.	Estudios teóricos ^a , FMPS, subescala del EDI.
2	Venezuela	Lic. en Psicología, Mgtr. en Psicología.	Perfeccionismo académico
3	Chile	Médico cirujano, Especialista en Psiquiatría.	Estudios teóricos
4	España	Lic. en Filosofía, Lic. en Psicología, Dra. en Psicología, Especialista en Psicología Clínica.	FMPS
5	España	Lic. en Psicología, Especialista en Psicología Clínica, Lic. en Filosofía.	FMPS
6	España	Médico, Especialista en Psiquiatría, PhD en Medicina, Mtr. en Gestión Hospitalaria.	CAPS, PSPS, subescala del EDI-2.
7	España	Lic. en Educación Física, Dr. en Ciencias de la Educación.	IPI
8	Venezuela	Lic. en Educación, Mgtr. en Psicología, Dra. en Ciencias Sociales y Humanidades.	FMPS
9	España	Lic. en Educación Física, Mtr. en Alto Rendimiento Deportivo, Dr. en Pedagogía.	IPI

(continúa)

(viene)

Núm. Experto	País	Formación académica	Marco de investigación
10	México	Lic. en Psicología, Dra. en Psicología.	Estudios teóricos, FMPS, subescala del EDI.
11	Argentina	Lic. en Eugenesia Integral y Humanismo, Lic. en Psicología, Dra. en Psicología.	Estudios teóricos
12	España	Lic. en Psicología, Dr. en Psicología.	IPI
13	España	Psicóloga, Mtr. en Psicopatología Clínica de Adultos, Mtr. en Terapia de Conducta, PhD en Psicología.	FMPS, HMPS, subescala del EDI-2.
14	España	—	FMPS, HMPS, subescala del EDI-2.
15	España	Lic. en Psicología, Dr. en Psicología.	IPI
16	España	Lic. en Psicología, Maestro especialista en Pedagogía Terapéutica, Dr. en Psicología.	IPI
17	España	—	CAPS, subescala del EDI-2.
18	España	Lic. en Educación Física, Dr. en Ciencias de la Educación Física y Deportes.	IPI
19	España	Lic. en Ciencias de la Actividad Física y el Deporte, Dr. en Educación Física.	FMPS, subescala del EDI-2.
20	Venezuela	Lic. en Psicología.	Perfeccionismo académico
21	España	Lic. en Psicología, Mtr. de Investigación en Psicología de la Salud, PhD en Psicología.	IPI
22	España	Lic. en Psicología, Dr. en Psicología Clínica y de la Salud.	CAPS
23	Argentina	Lic. en Psicología.	FMPS
24	Venezuela	tsu en Terapia Ocupacional, Lic. en Psicología.	Perfeccionismo académico
25	España	Lic. en Psicología, Mtr. en Psicología Clínica y de la Salud, Dra. en Psicología de la Salud.	CAPS
26	Argentina	Profa. de Enseñanza Media y Superior en Ciencias de la Educación, Lic. en Psicología, Dra. en Psicología.	FMPS
27	Argentina	Lic. en Psicología, Posgrado en Teoría y Práctica Clínica de Terapia Cognitiva, Especialista en Psicología Clínica de niños, adolescentes y familias.	APS-R
28	Argentina	Lic. en Psicología, Residencia completa de Salud Mental, Programa de Actualización en Terapia Cognitiva, Doctoranda en Psicología.	Estudios teóricos, APS-R
29	España	Médico psiquiatra, Dra. en Medicina.	FMPS, HMPS, subescala del EDI-2.
30	México	Lic. en Psicología, Maestra en Psicopatología Infanto-Juvenil, Doctoranda en Psicología.	Estudios teóricos, FMPS, subescala del EDI.
31	Paraguay	Psicóloga, Maestría - Special Education, Maestría - Counseling Psychology, Doctoranda.	CAPS

Nota. Las celdillas con una raya (—) no se pudieron llenar porque no se obtuvieron/no se informaron los datos. Instrumentos de medición: FMPS = Frost Multidimensional Perfectionism Scale (Frost et al., 1990; versiones españolas de Carrasco et al., 2010 y de Gelabert et al., 2011; versión mexicana de Franco et al., 2010; versión venezolana de Prieto y Vento, 2012, como se citó en Elvira, 2015); EDI = Eating Disorder Inventory (Garner et al., 1983; adaptado para población española por Guimerá et al., 1987; en tanto que su estructura factorial se corroboró en la validación con población mexicana: Álvarez y Franco, 2001; Álvarez et al., 2000; Franco et al., 2004; Mancilla et al., 2003); CAPS = Child-Adolescent Perfectionism Scale (Flett et al., 2000; versión española de Castro et al., 2004); PSPS = Perfectionistic Self-Presentation Scale (Hewitt et al., 2003; versión española de Castro et al., 2004); EDI-2 = Eating Disorder Inventory-2 (Garner, 1991, 1998); IPI = Inventario de Perfeccionismo Infantil (Lozano et al., 2012); HMPS = Hewitt Multidimensional Perfectionism Scale (Hewitt y Flett, 1991, 2004; versiones españolas de Rodríguez-Campayo et al., 2009 y de Gelabert et al., 2011); APS-R = Almost Perfect Scale Revised (Slaney et al., 1996; Slaney et al., 2001; versión argentina de Arana et al., 2009). ^a Se considera estudio teórico “todo aquél que presenta avances teóricos, estudios de revisión, actualización, comparación y análisis crítico de teorías o modelos en un determinado campo” (Montero y León, 2007, p. 849). Elaboración propia.

Instrumento objeto de validación

La Escala de Perfeccionismo Infantil construida por Oros (2003) es un instrumento que permite evaluar el perfeccionismo en niños. Consta de 16 ítems que se agrupan en dos dimensiones: (a) Autodemandas y (b) Reacciones ante el fracaso; cada una cuenta con ocho ítems, a los que se responde mediante una escala tipo Likert de tres puntos (1 = no o no lo pienso, 2 = a veces o lo pienso algunas veces, 3 = sí o lo pienso muchas veces).

El cuestionario fue estudiado en una muestra de 583 niños argentinos escolarizados (309 niñas y 274 varones), comprendidos en un rango de edad de 8 a 13 años ($M = 11,03$), de nivel socioeconómico medio.

La extracción de los factores siguió, en primer lugar, un criterio teórico bajo el cual se hipotetizó que los reactivos se agruparían según el modelo de Hewitt y Flett (1991), es decir, que los pensamientos perfeccionistas orientados hacia sí mismo se distinguirían del perfeccionismo socialmente prescripto. Sin embargo, los ítems que reflejaban cogniciones perfeccionistas se agruparon en un factor que podría denominarse “Autodemandas”, mientras que ciertas emociones y actitudes asociadas al fracaso de estas autodemandas se agruparon en un segundo factor que podría ser denominado como “Reacción ante el fracaso”.

La dimensión de las Autodemandas, que expresa un perfeccionismo orientado hacia sí mismo, evalúa la tendencia de los niños a exigirse la perfección y a evitar constantemente los errores y equivocaciones. Este factor incluye creencias irracionales del tipo: “Necesito ser el mejor” y “Debo ganar siempre”. La dimensión de Reacciones ante el fracaso refleja las emociones y actitudes asociadas al fracaso de las autodemandas. La forma venezolana de la escala incluye en este factor enunciados como los siguientes: “Me critico mucho” o “Me cuesta mucho aceptar que me equivoqué”.

Debido a que el segundo factor incluía reacciones emotivas muy diferentes, la autora decidió realizar una refactorización, tras lo cual encontró que las emociones asociadas al fracaso se podían clasificar en tres grupos: (a) Culpa y remordimiento, (b) Malestar y (c) Autocastigo.

Se hallaron niveles de consistencia interna satisfactorios en el medio argentino, tanto en relación con la escala total ($\alpha = 0,83$) como con cada subescala: Autodemandas ($\alpha = 0,82$) y Reacciones ante el fracaso ($\alpha = 0,70$).

Materiales y procedimiento

Lo habitual es que el juicio sobre la correspondencia de los reactivos con las dimensiones de la prueba se lleve a cabo por medio de algún procedimiento estructurado (Martínez, 1996). En este caso, los expertos recibieron una comunicación electrónica en la que se les expuso el propósito del estudio, es decir, proveer a la escala de validez de contenido. Como archivo adjunto recibían un formato que debían descargar. En él se encontraban, sin ningún orden específico, los ítems que conforman la versión venezolana de la escala. La tarea del experto era juzgar qué evaluaba cada reactivo, seleccionando la dimensión a la que suponía que pertenecía cada ítem. Para esto se suministraba como apoyo la conceptualización teórica de la variable examinada por el instrumento y la de sus facetas o dimensiones. Tomando en cuenta que los contactados, en su mayoría, eran extranjeros, se les aclaró el significado de dos términos que forman parte de la redacción de los reactivos (viz. “el salón”¹ y “rabia”²), los cuales podían generar confusión al tener otra connotación en sus países de origen. Los expertos tardaban alrededor de tres días en responder.

Aunque normalmente en la validación de contenido las decisiones son más cualitativas que cuantitativas (Martínez, 1996), existen diversos indicadores numéricos para cuantificar el grado de acuerdo entre los expertos (Anastasi y Urbina, 1998; Hogan, 2003; Kaplan y Saccuzzo, 2006; Muñiz, 1992; Pedrosa et al., 2013). Se utilizó el Índice de Validez Factorial (FVI, por sus siglas en inglés), el cual aporta información acerca del grado en que los expertos asocian cada ítem con los constructos que se pretenden medir, aportando así una “cuantificación preliminar de la validez factorial” (Rubio et al., 2003, p. 98). Se trata de un método relativamente reciente. Un ejemplo de su aplicación puede ser consultado en el interesante trabajo de Yang y Chan (2008), acerca del diseño de páginas web para el aprendizaje de idiomas. Para calcular el FVI de cada ítem, se dividió el número de expertos que asociaron correctamente el ítem con su dimensión entre los expertos totales. Este mismo procedimiento, tomando la media del FVI a lo largo de los diferentes ítems, se empleó para calcular el FVI del instrumento total. A la hora de interpretar el

1 El salón: la clase, el curso, el aula.

2 Rabia: ira, enojo, enfado

resultado, Rubio et al. (2003) proponen alcanzar un valor de 0,80 para considerar tanto el ítem como la escala adecuados.

Resultados

En general, se puede observar (véase la tabla 2) que, con excepción de los reactivos 12 y 13 que obtuvieron FVI muy bajos (0,06 y 0,10, respectivamente) los ítems presentaron FVI que van desde 0,81 a 1, y el FVI de la escala total fue de 0,85, el cual resulta aceptable al estar por encima de 0,80. En este sentido, seis ítems (2, 3, 5, 7, 8 y 16) tienen un FVI de 1, cuatro ítems (1, 4, 6 y 9) muestran un FVI de 0,97, y los demás ítems (15, 14, 10 y 11, en ese orden) exhiben un FVI que oscila entre 0,81 y 0,94.

Así mismo, se aprecia que todos los expertos consiguieron asignar correctamente cinco ítems (2, 3, 5, 7 y 8) a la dimensión Autodemandas ($FVI = 1$), mientras que 30 expertos asociaron de forma correcta tres ítems (4, 6 y 10) a dicha dimensión ($FVI = 0,97$), para un FVI promedio igualmente aceptable de 0,99.

Dentro de la dimensión Reacciones ante el fracaso, el FVI promedio de emparejamiento con las subdimensiones correspondientes fue de 0,71, el cual cae por debajo del valor de 0,80 requerido.

Así, 28 expertos lograron parear acertadamente el ítem 10 con la subdimensión Culpa y remordimiento

($FVI = 0,90$), por su parte, 25 expertos clasificaron correctamente el ítem 15 en dicha subdimensión ($FVI = 0,81$), para un FVI promedio de 0,86.

También se observa que 30 expertos acertaron al señalar que el ítem 9 pertenecía a la subdimensión Malestar ($FVI = 0,97$) y 29 expertos consideraron que el ítem 11 era propio de tal subdimensión ($FVI = 0,94$), para un FVI promedio de 0,96.

Finalmente, se puede apreciar que la totalidad de los expertos emparejó correctamente el ítem 16 con la subdimensión Autocastigo ($FVI = 1$), y 27 expertos vincularon de forma adecuada el ítem 14 con dicha subdimensión ($FVI = 0,87$). Sin embargo, para el ítem 13 el FVI fue de 0,10, lo que quiere decir que solo tres expertos estimaron el contenido de ese ítem como una forma de Autocastigo; de hecho, la mayoría (25 expertos) lo relacionó con Malestar y otros tres expertos lo juzgaron como perteneciente a Culpa y remordimiento. De igual manera, apenas dos expertos catalogaron el ítem 12 dentro de Autocastigo ($FVI = 0,06$); más bien, la mayoría (16 expertos) lo encasilló como parte de la subdimensión Culpa y remordimiento, otros siete expertos lo asociaron a la dimensión Autodemandas y, por último, seis a la subdimensión Malestar, para un FVI medio de 0,51 (por debajo del requisito de 0,80) en la subdimensión de Autocastigo.

Tabla 2

Resultados del estudio de validez de contenido a través del juicio de expertos

Ítems	AD		CR		M		AC		FVI
	f	%	f	%	f	%	f	%	
2. Tengo que ser el mejor alumno	31	100	0	0	0	0	0	0	1
3. Debo ganar siempre	31	100	0	0	0	0	0	0	1
5. Mis trabajos deben ser mejores que los de mis compañeros	31	100	0	0	0	0	0	0	1
7. Debo ser el mejor del salón	31	100	0	0	0	0	0	0	1
8. Debo ser el primero en terminar las tareas	31	100	0	0	0	0	0	0	1
1. Necesito ser el mejor	30	96,8	0	0	1	3,2	0	0	0,97
4. No puedo cometer errores	30	96,8	0	0	1	3,2	0	0	0,97
6. No debo perder cuando juego con mis amigos	30	96,8	0	0	1	3,2	0	0	0,97
10. Pienso mucho en qué me equivoqué	0	0	28	90,3	3	9,7	0	0	0,90
15. Me siento culpable cuando cometo algún error	0	0	25	80,6	4	13	2	6,4	0,81
12. Me cuesta mucho aceptar que me equivoqué	7	22,6	16	51,7	6	19,3	2	6,4	0,06

(continúa)

(viene)

Ítems	AD		CR		M		AC		FVI
	f	%	f	%	f	%	f	%	
9. Cuando pierdo me siento mal	0	0	1	3,2	30	96,8	0	0	0,97
11. Me siento muy mal cuando pierdo en algún juego o deporte	1	3,2	1	3,2	29	93,6	0	0	0,94
13. Me da rabia cuando no logro lo que quiero	0	0	3	9,7	25	80,6	3	9,7	0,10
16. Me insulto cada vez que cometo un error	0	0	0	0	0	0	31	100	1
14. Me critico mucho	1	3,2	3	9,7	0	0	27	87,1	0,87
	Total								0,85

Nota. Los ítems en negritas no tuvieron un FVI adecuado y pertenecen a la subdimensión Autocastigo. AD = Autodemandas; CR = Culpa y remordimiento; M = Malestar; AC = Autocastigo. Elaboración propia.

Discusión

Con frecuencia, la determinación de la validez de contenido la realizan jueces expertos (Kaplan y Saccuzzo, 2006). Por lo común, el procedimiento de evaluación consiste en el emparejamiento de los ítems con los contenidos que han de ser evaluados (Abad et al., 2011). Los ítems de la versión venezolana de la Escala de Perfeccionismo Infantil fueron sometidos a la valoración de expertos en el área, a fin de garantizar que tuvieran relación directa con lo que, en teoría, afirmaban medir. La tarea de los expertos era determinar a qué dimensión correspondía cada ítem, conociendo el modo como se conceptualizaba la variable perfeccionismo en esta investigación. Aunque, en sentido estricto, la validez de contenido no puede expresarse por un coeficiente de validez (Magnusson, 1975), se han propuesto algunos índices cuantitativos para resumir las decisiones de los jueces sobre los ítems (Hogan, 2003; Martínez, 1996; Pedrosa et al., 2013). Aquí se empleó el Índice de Validez Factorial (FVI) planteado por Rubio et al. (2003).

Si bien, tal como afirma Muñoz (1992), la validez por expertos sigue un camino más indirecto y falible, siguiendo lo planteado por los autores especializados en esta metodología (e.g., Davis, 1992; Tornimbeni et al., 2004), se seleccionaron probados expertos en la temática objeto de investigación (i.e., el perfeccionismo). De igual manera, aunque no hay un acuerdo unánime para la determinación del número de expertos necesarios, los autores recomiendan un mínimo de tres (Delgado-Rico et al., 2012; Lynn, 1986), elevando el rango hasta 30 (Gable y Wolf, 1993; Mikulic, 2007; Rubio et al., 2003; Varela et al., 2012; Walz et al., 1991). Al tomar en cuenta dichas precisiones, se siguió un procedimiento sistematizado de consulta a un amplio número de expertos.

En la consulta con los expertos en la variable, hubo un menor porcentaje de acuerdo al clasificar los reactivos de la dimensión Reacción ante el fracaso que al hacer lo propio con los de la dimensión Autodemandas. Es probable que la situación guarde relación con el hecho de que dicha dimensión esté integrada por ítems que refieren reacciones emocionales, las cuales se ven influenciadas en alguna medida por diversos elementos socioculturales (Palmero et al., 2002; Reeve, 1994).

El contenido original del reactivo 12 (Me cuesta perdonarme cuando me equivoco) fue incluido como un tipo de Autocastigo. La perspectiva ofrecida por la autora de la escala argentina es que sentirse culpable y recordar los errores personales es muy distinto a no perdonarse el haberlos cometido, esto último sería, en efecto, una especie de autocastigo (L. B. Oros, comunicación personal, 19 de octubre de 2015). Por su parte, la forma venezolana de dicho ítem (“Me cuesta mucho aceptar que me equivoqué”) fue mayormente vinculada por los expertos con la dimensión Culpa y remordimiento (52%), y aunque fue muy disputado su dominio (siendo el ítem que registró un menor porcentaje de acuerdo), solo unos pocos jueces (6%) lo percibieron como una muestra de Autocastigo. A esto se debe agregar que, en el factorial exploratorio argentino, la segunda saturación en orden de importancia (0,28) del elemento que se discute recayó en la dimensión Culpa y remordimiento. Faltaría ver a qué dimensión se adhiere este ítem dentro de la ordenación factorial que la escala muestre en el caso venezolano.

La versión argentina del ítem 13 (Me enojo cuando no logro lo que quiero) se asoció a los ítems de Autocastigo. De acuerdo con L. B. Oros (comunicación personal, 19 de octubre de 2015), quizá esto revele que ese enojo no se refiere a un malestar

general, sino a un marcado enojo contra sí mismo. Ciertamente, algunos individuos necesitan sentirse verdaderamente mal antes de pasar a sentirse mejor; en ese ínterin se autorreprochan, dirigen la ira hacia sí mismos o emplean cualquier otra forma de autocastigo (Londoño et al., 2006). Sin embargo, la versión venezolana de ese reactivo (Me da rabia cuando no logro lo que quiero) fue considerada por la mayoría de los especialistas en el perfeccionismo (80,6%) como parte de la dimensión Malestar, mientras que solo una minoría (9,7%) valoró el ítem 13 como una manera de autocastigarse. Dado que algunos expertos (9,7%) también vincularon el ítem a la dimensión Culpa y remordimiento, y si se tiene en cuenta que su segunda carga de importancia en el análisis factorial argentino (0,32) pertenece a la dimensión Malestar, no puede sino concluirse que dicho reactivo refiere a una compleja experiencia emocional que probablemente esté modulada por diferencias culturales (i.e., quizá la rabia no figura como una forma típica de autocastigo en un niño venezolano).

La valoración cualitativa realizada por los expertos en el campo del perfeccionismo a la versión venezolana de la escala merece dos observaciones críticas. En primer lugar, el hecho de que varios jueces hayan calificado cada reactivo en términos de su relevancia con el contenido no agota el proceso de validación, puesto que analizar la validez de contenido también implica valorar aspectos como la utilidad y la importancia de cada ítem (Sireci, 1998a).

En segundo lugar, es obvio que los expertos tenían conocimiento del rasgo que se pretendía medir y, como parte de las instrucciones de la tarea que les fue asignada, recibieron información sobre los aspectos considerados por la teoría implicada. Según Pedrosa et al. (2013), con esto último existe el riesgo de “dirigir” la valoración, lo cual puede provocar un incremento artificial de las tasas de acuerdo entre los expertos y alterar así la información real acerca del instrumento. Sin embargo, a pesar de la importancia de este posible sesgo, son escasas las estrategias desarrolladas para evitarlo y, tal como nos lo recuerda Sireci (2007), resulta complicado imaginar que un grupo de expertos evalúe un determinado número de ítems sin que puedan conocer —al menos intuitivamente— lo que pretenden medir.

Como ya se había apuntado, en la práctica los distintos tipos de validez tienden a complementarse entre sí. En este sentido, los procedimientos requeridos para asegurar la validez de contenido apoyan, a su vez, la validez de constructo (Brown, 1980; Cronbach

y Meehl, 1955; Nunnally y Bernstein, 1995); de hecho, Ebel (1956) llegó a considerar la primera como la base de la segunda. El método que se empleó para estimar la validez de contenido (i.e., el FVI desarrollado por Rubio et al., 2003) la relaciona explícitamente con la validez de constructo, de modo que la información proveniente del juicio de expertos podrá ser atendida a la luz de un próximo estudio sobre la validez de la escala por vía factorial en el medio venezolano.

Referencias

- Abad, F., Olea, J., Ponsoda, V. y García, C. (2011). *Medición en ciencias sociales y de la salud*. Madrid: Síntesis.
- Álvarez, D. y Franco, K. (2001). *Validación del Eating Disorder Inventory en población mexicana* (Tesis de licenciatura inédita). Universidad Nacional Autónoma de México, Ciudad de México, México.
- Álvarez, D., Franco, K., Mancilla, J., Álvarez, G. y López, X. (2000). Factores predictores de la sintomatología de los trastornos alimentarios. *Psicología Contemporánea*, 7(1), 26-35.
- Anastasi, A. y Urbina, S. (1998). *Test psicológicos*. (7.ª ed.). Ciudad de México: Prentice Hall.
- Arana, F., Keegan, E. y Rutsztein, G. (2009). Adaptación de una medida multidimensional de perfeccionismo: la Almost Perfect Scale-Revised (APS-R). Un estudio preliminar sobre sus propiedades psicométricas en una muestra de estudiantes universitarios argentinos. *Evaluar*, 9, 35-53. Recuperado de <http://revistas.unc.edu.ar/index.php/revaluar/articulo/view/463>
- Brown, F. (1980). *Principios de la medición en psicología y educación* (4.ª ed.). Ciudad de México: El Manual Moderno.
- Carrasco, A., Belloch, A. y Perpiñá, C. (2010). La evaluación del perfeccionismo: utilidad de la Escala Multidimensional de Perfeccionismo en población española. *Análisis y Modificación de Conducta*, 36(153), 49-65. Recuperado de <http://www.uhu.es/publicaciones/ojs/index.php/amc/article/view/1225>
- Carretero-Dios, H. y Pérez, C. (2007). Standards for the development and review of instrumental studies: Considerations about test selection in psychological research. *International Journal of Clinical and Health Psychology*, 7(1), 863-882. Recuperado de http://www.aepc.es/ijchp/articulos_pdf/ijchp-257.pdf
- Castro, J., Gila, A., Gual, P., Lahortiga, F., Saura, B. y Toro, J. (2004). Perfectionism dimensions in children and adolescents with anorexia nervosa. *The Journal of Ad-*

- Adolescent Health: Official Publication of the Society for Adolescent Medicine*, 35(5), 392-398. doi: 10.1016/j.jadohealth.2003.11.094
- Cohen, R. y Swerdlik, M. (2001). *Pruebas y evaluación psicológica: introducción a las pruebas y a la medición*. (4.ª ed.). Ciudad de México: McGraw-Hill.
- Cronbach, L. y Meehl, P. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281-302. doi: 10.1037/h0040957
- Davis, L. (1992). Instrument review: Getting the most from your panel of experts. *Applied Nursing Research*, 5(4), 194-197. doi: 10.1016/S0897-1897(05)80008-4
- Delgado-Rico, E., Carretero-Dios, H. y Ruch, W. (2012). Content validity evidences in test development: An applied perspective. *International Journal of Clinical and Health Psychology*, 12(3), 449-460. Recuperado de http://www.aepc.es/ijchp/articulos_pdf/ijchp-421.pdf
- Dembo, M. (1983). Dos estrategias de medición en psicología. *Revista Psicología*, 10(1-2), 41-57.
- Ebel, R. (1956). Obtaining and reporting evidence for content validity. *Educational and Psychological Measurement*, 16(3), 269-282. doi: 10.1177/001316445601600301
- Elvira, M. (2015). *Perfil familiar y psicológico como predictores del estrés académico: un modelo de ruta* (Tesis doctoral inédita). Universidad Simón Bolívar, Caracas, Venezuela.
- Flett, G., Hewitt, P., Boucher, D., Davidson, L. y Munro, I. (2000). *The Child-Adolescent Perfectionism Scale: Development, validation, and association with adjustment*. Manuscrito inédito, University of British Columbia, Vancouver, Canada.
- Franco, K., Mancilla, J., Álvarez, G., Vázquez, R. y López, X. (2004). *Internal consistency and solution factor of the Eating Disorders Inventory (EDI) in Mexican males*. Sesión de cartel presentada en la International Conference on Eating Disorders (ICED) de la Academy for Eating Disorders (AED), Orlando, Florida, Estados Unidos, mayo de 2004.
- Franco, K., Mancilla-Díaz, J., Vázquez, R., Álvarez, G. y López, X. (2010). Estructura factorial y consistencia interna de la Escala Multidimensional de Perfeccionismo. *Revista Mexicana de Psicología*, 27(2), 143-149. Recuperado de <http://www.redalyc.org/articulo.oa?id=243016324003>
- Frost, R., Marten, P., Lahart, C. y Rosenblate, R. (1990). The dimensions of perfectionism. *Cognitive Therapy and Research*, 14(5), 449-468. doi: 10.1007/BF01172967
- Gable, R. y Wolf, J. (1993). *Instrument development in the affective domain: Measuring attitudes and values in corporate and school settings* (2.ª ed.). Nueva York: Kluwer Academic.
- Garner, D. (1991). *Eating Disorder Inventory-2: Professional Manual*. Odessa: Psychological Assessment Resources.
- Garner, D. (1998). *Inventario de trastornos de la conducta alimentaria (EDI-2)*. Madrid: TEA.
- Garner, D., Olmstead, M. y Polivy, J. (1983). Development and validation of a multidimensional Eating Disorder Inventory for anorexia nervosa and bulimia. *International Journal of Eating Disorders*, 2(2), 15-34. doi: 10.1002/1098-108X(198321)2:23.0.CO;2-6
- Gelabert, E., García-Esteve, L., Martín-Santos, R., Gutiérrez, F., Torres, A. y Subirà, S. (2011). Psychometric properties of the spanish version of the Frost Multidimensional Perfectionism Scale in women. *Psicothema*, 23(1), 133-139. Recuperado de <http://www.psicothema.com/psicothema.asp?id=3861>
- Guimerá, E., Querol, E., y Torrubia, R. (1987). Adaptación española del Eating Disorder Inventory (EDI) en una muestra de pacientes anoréxicas. *Anales de Psiquiatría*, 3, 185-190.
- Hewitt, P. y Flett, G. (1991). Perfectionism in the self and social contexts: Conceptualization, assessment, and association with psychopathology. *Journal of Personality and Social Psychology*, 60(3), 456-470. doi: 10.1037/0022-3514.60.3.456
- Hewitt, P. y Flett, G. (2004). *Multidimensional Perfectionism Scale (MPS): Technical manual*. Toronto: Multi-Health Systems.
- Hewitt, P., Flett, G., Sherry, S., Habke, M., Parkin, M., Lam, R., ..., Stein, M. (2003). The interpersonal expression of perfection: Perfectionistic self-presentation and psychological distress. *Journal of Personality and Social Psychology*, 84(6), 1303-1325. doi: 10.1037/0022-3514.84.6.1303
- Hogan, T. (2003). *Pruebas psicológicas: una introducción práctica*. Ciudad de México: El Manual Moderno.
- Kane, M. (2006). Content-related validity evidence in test development. En S. Downing y T. Haladyna (Eds.), *Handbook of test development* (pp. 131-153). Mahwah: Lawrence Erlbaum Associates.
- Kane, M. (2009). Validating the interpretations and uses of test scores. En R. Lissitz (Ed.), *The concept of validity: Revisions, new directions and applications* (pp. 39-64). Charlotte: Information Age Publishing.

- Kaplan, R. y Saccuzzo, D. (2006). *Pruebas psicológicas: principios, aplicaciones y temas* (6.ª ed.). Ciudad de México: Thomson.
- Londoño, N., Henao, G., Puerta, I., Posada, S., Arango, D. y Aguirre-Acevedo, D. (2006). Propiedades psicométricas y validación de la Escala de Estrategias de Coping Modificada (EEC-M) en una muestra colombiana. *Universitas Psychologica*, 5(2), 327-349. Recuperado de <http://www.redalyc.org/articulo.oa?id=64750210>
- Lozano, L., García, E., Martín, M. y Lozano, L. (2012). Desarrollo y validación del Inventario de Perfeccionismo Infantil (IPI). *Psicothema*, 24(1), 149-155. Recuperado de <http://www.psicothema.com/psicothema.asp?id=3992>
- Lozano, L., Llanos, A., Lozano, L. y Valor-Segura, I. (2014). El efecto del perfeccionismo infantil en la aparición de sintomatología internalizada. En M. Ramiro y T. Ramiro-Sánchez (Comps.), *vii Congreso Internacional y xii Nacional de Psicología Clínica* (p. 276). Sevilla: Asociación Española de Psicología Conductual (AEPCC).
- Lozano, L., Valor-Segura, I., Llanos, A. y Lozano, L. (2014). Efectos del perfeccionismo infantil sobre la inteligencia y el rendimiento académico. En A. Romero, T. Ramiro-Sánchez y M. Paz (Coords.), *Libro de Actas del II Congreso Internacional de Ciencias de la Educación y del Desarrollo* (p. 178). Granada: Asociación Española de Psicología Conductual (AEPCC).
- Lynn, M. (1986). Determination and quantification of content validity. *Nursing Research*, 35(6), 382-385. doi: 10.1097/00006199-198611000-00017
- Magnusson, D. (1975). *Teoría de los tests*. Ciudad de México: Trillas.
- Mancilla, J., Franco, K., Álvarez, G. y Vázquez, R. (2003). Evaluation of the psychometric properties of the Mexican version of the Eating Disorders Inventory. *Thomson Psicología*, 1(2), 167-176.
- Martínez, R. (1996). *Psicometría: teoría de los tests psicológicos y educativos*. Madrid: Síntesis.
- Matesanz, A. (1997). *Evaluación estructurada de la personalidad*. Madrid: Pirámide.
- Mikulic, I. (2007). *Construcción y adaptación de pruebas psicológicas*. Manuscrito inédito, Universidad de Buenos Aires, Argentina.
- Montero, I. y León, O. (2005). Sistema de clasificación del método en los informes de investigación en Psicología. *International Journal of Clinical and Health Psychology*, 5(1), 115-127. Recuperado de http://www.aepc.es/ijchp/articulos_pdf/ijchp-136.pdf
- Montero, I. y León, O. (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/ijchp/articulos_pdf/ijchp-256.pdf
- Moreno, R., Martínez, R. y Muñiz, J. (2004). Directrices para la construcción de ítems de elección múltiple. *Psicothema*, 16(3), 490-497. Recuperado de <http://www.psicothema.com/psicothema.asp?id=3023>
- Muñiz, J. (1992). *Teoría clásica de los tests*. Madrid: Pirámide.
- Nunnally, J. y Bernstein, I. (1995). *Teoría psicométrica* (3.ª ed.). Ciudad de México: McGraw-Hill.
- Oros, L. (2003). Medición del perfeccionismo infantil: desarrollo y validación de una escala para niños de 8 a 13 años de edad. *Revista Iberoamericana de Diagnóstico y Evaluación Psicológica*, 16(2), 99-112. Recuperado de http://www.aidep.org/03_ridep/R16/R166.pdf
- Palmero, F., Fernández-Abascal, E., Martínez, F. y Chóliz, M. (2002). *Psicología de la Motivación y la Emoción*. Madrid: McGraw-Hill.
- Pedrosa, I., Suárez-Álvarez, J. y García-Cueto, E. (2013). Evidencias sobre la validez de contenido: avances teóricos y métodos para su estimación. *Acción Psicológica*, 10(2), 3-18. doi: 10.5944/ap.10.2.11820
- Reeve, J. (1994). *Motivación y emoción*. Madrid: McGraw-Hill.
- Rodríguez-Campayo, M., Rojo-Moreno, L., Ortega-Toro, E. y Sepúlveda-García, A. (2009). Adaptación de la Escala Multidimensional de Perfeccionismo a estudiantes universitarios españoles. *Ansiedad y Estrés*, 15(1), 13-27.
- Rubio, D., Berg-Weger, M., Tebb, S., Lee, E. y Rauch, S. (2003). Objectifying content validity: Conducting a content validity study in social work research. *Social Work Research*, 27(2), 94-104. doi: 10.1093/swr/27.2.94
- Sireci, S. (1998a). The construct of content validity. *Social Indicators Research*, 45(1), 83-117. doi: 10.1023/A:1006985528729
- Sireci, S. (1998b). Gathering and analyzing content validity data. *Educational Assessment*, 5(4), 299-321. doi: 10.1207/s15326977ea0504_2
- Sireci, S. (2003). Validity content. En R. Fernández-Ballesteros (Ed.), *Encyclopedia of psychological assessment* (pp. 1076-1078). Londres: SAGE Publications.
- Sireci, S. (2007). On validity theory and test validation. *Educational Researcher*, 36(8), 477-481. doi: 10.3102/0013189X07311609

- Sireci, S. (2009). Packing and unpacking sources of validity evidence. En R. Lissitz (Ed.), *The concept of validity: Revisions, new directions and applications* (pp. 19-37). Charlotte: Information Age Publishing.
- Slaney, R., Mobley, M., Trippi, J., Ashby, J. y Johnson, D. (1996). *Almost Perfect Scale-Revised*. Manuscrito inédito, Pennsylvania State University, Estados Unidos.
- Slaney, R., Rice, K., Mobley, M., Trippi, J. y Ashby, J. (2001). The revised Almost Perfect Scale. *Measurement and Evaluation in Counseling and Development*, 34(3), 130-145.
- Tornimbeni, S., Pérez, E., Olaz, F. y Fernández, A. (2004). *Introducción a los tests psicológicos* (3.ª ed.). Córdoba: Brujas.
- Varela Ruiz, M., Díaz-Bravo, L. y García-Durán R. (2012). Descripción y usos de la técnica Delphi en investigaciones del área de la salud. *Investigación en Educación Médica*, 1(2), 90-95. Recuperado de <http://riem.fac-med.unam.mx/node/204>
- Walz, C., Strickland, O. y Lenz, E. (1991). *Measurement in nursing research*. (2.ª ed.). Filadelfia: F. A. Davis.
- Yang, Y. y Chan, C. (2008). Comprehensive evaluation criteria for English learning websites using expert validity surveys. *Computers & Education*, 51(1), 403-422. doi:10.1016/j.compedu.2007.05.011