

MSLQ-UY, validación con estudiantes universitarios uruguayos

MSLQ-UY, validation with Uruguayan university students

Karina Curione *¹, Virginia Gründler², Lucía Píriz³, Juan Antonio Huertas⁴

1 - Instituto de Fundamentos y Métodos en Psicología, Facultad de Psicología, Universidad de la República, Uruguay.

2 - Administración Nacional de Educación Pública - Consejo de Educación Secundaria, Uruguay.

3 - Facultad de Psicología, Universidad de la República, Uruguay.

4 - Facultad de Psicología, Universidad Autónoma de Madrid, España.

Introducción
Método
Resultados
Discusión
Conclusiones
Referencias

Recibido: 23/04/2017 Revisado: 17/05/2017 Aceptado: 29/05/2017

Resumen

Se presenta la traducción, adaptación y validación al español de la versión abreviada del Motivated Strategies for Learning Questionnaire (MSLQ), desarrollado por Pintrich y De Groot (1990) para población uruguaya universitaria. El MSLQ fue construido desde el marco socio-cognitivo de la motivación y la autorregulación. El instrumento contiene 44 ítems que se responden con una escala tipo Likert de 1 a 7 y se divide en dos bloques denominados *motivación* y *estrategias de aprendizaje autorregulado* (EAA). La muestra incluyó 318 estudiantes de la Facultad de Psicología de la Universidad de la República (Uruguay). Los resultados evidencian que el bloque de *motivación* ($\alpha = .76$) y el de *estrategias de aprendizaje autorregulado* ($\alpha = .75$) poseen una consistencia interna aceptable para ser utilizados con fines de investigación. El análisis factorial exploratorio permitió confirmar la estructura original de cinco factores, aunque se plantea una reorganización del bloque *estrategias de aprendizaje autorregulado*.

Palabras clave: MSLQ, validación, motivación, aprendizaje autorregulado, Psicología

Abstract

The current study translates, adapts, and validates for Uruguayan university students the abbreviated Spanish version of the Motivated Strategies for Learning Questionnaire (MSLQ) developed by Pintrich and De Groot (1990). The MSLQ was created within a motivational and self-regulated socio-cognitive framework. This questionnaire includes 44 items, and the response choices are rated on a 7-point Likert-type scale. MSLQ is divided into two sections: *motivation* and *self-regulated learning strategies* (SRLS). The sample included 318 students total who belong the Psychology department of Universidad de la República (Uruguay). Results have shown that both *motivation* ($\alpha = .76$) and *self-regulated learning strategies* ($\alpha = .75$) sections have acceptable internal consistency to be used in research studies. Although the exploratory factor analysis confirms the original five-factor structure, we suggest a reorganization of the *self-regulated learning strategies* section.

Keywords: MSLQ, validation, motivation, self-regulated learning, Psychology

*Correspondencia a: Karina Curione. Dirección postal: Tristán Narvaja 1674, CP 11200, Montevideo, Uruguay. Teléfono: (+598) 24008555. kcurione@psico.edu.uy

Cómo citar este artículo: Curione, K., Gründler, V., Píriz, L., & Huertas, J. A. (2017). MSLQ-UY, validación con estudiantes universitarios uruguayos. *Revista Evaluar*, 17(02), 1-17. Recuperado de <https://revistas.unc.edu.ar/index.php/revaluar>

Nota del autor: Este trabajo forma parte de la tesis de doctorado de la primera autora. Doctorado en Psicología, Universidad Católica del Uruguay.

Introducción

Uno de los principales antecedentes en el estudio de la motivación, la autorregulación y el rendimiento académico a nivel internacional son los diversos estudios realizados desde mediados de la década de 1980 por Paul Pintrich y colaboradores (Pintrich & De Groot, 1990; Pintrich, Smith, García, & McKeachie, 1991, 1993) los que condujeron a la creación del Motivated Strategies for Learning Questionnaire (MSLQ). El desarrollo de dicho instrumento comenzó en 1986 y supuso 10 años de trabajo de Pintrich y su equipo; durante ese tiempo, el instrumento fue utilizado en numerosos estudios correlacionales que involucraron a miles de estudiantes en sus muestras.

El MSLQ fue creado a partir del marco socio-cognitivo de la motivación y la autorregulación, el cual asume que la motivación está relacionada con la habilidad de autorregular las actividades de aprendizaje. Así, se considera que un estudiante necesita estar motivado para utilizar eficazmente las estrategias de aprendizaje (Pintrich & De Groot, 1990). El cuestionario se utiliza para evaluar la motivación y las estrategias de aprendizaje autorregulado que el estudiante emplea en un curso en particular.

Pintrich (2000) define el aprendizaje autorregulado como un proceso activo, constructivo, mediante el cual los aprendices eligen metas para su aprendizaje e intentan monitorear, regular y controlar su cognición, su motivación y sus conductas, guiados por las constricciones de sus propias metas y de los factores contextuales y ambientales. La investigación en el área ha permitido caracterizar al aprendiz autorregulado como alguien autónomo, reflexivo y eficiente, que posee las habilidades cognitivas y metacognitivas, así como las creencias motivacionales y las actitudes necesarias para comprender, monitorear y dirigir su propio aprendizaje (Wolters, 2003).

Las subescalas motivacionales del MSLQ se basan en tres constructos motivacionales: *expectativa*, *valor* y *afecto*. El componente *expectativa* hace referencia a las creencias de los estudiantes acerca de sus propias habilidades para desempeñarse en una tarea. El componente *valor* incluye las metas de los estudiantes y sus creencias acerca de la importancia e interés de la tarea. Por último, el componente *afectivo* remite a las reacciones emocionales de los estudiantes ante la tarea (Pintrich & De Groot, 1990).

El MSLQ en su versión abreviada (Pintrich & De Groot, 1990), objeto de la presente validación, contiene 44 ítems que se responden con una escala de tipo Likert cuyo rango de respuestas abarca de 1 a 7. El cuestionario se divide en dos partes, denominadas *motivación* y *estrategias de aprendizaje autorregulado* (EAA). Estas a su vez se dividen en subescalas. Motivación se divide en autoeficacia, valor intrínseco y ansiedad ante las pruebas; estrategias de aprendizaje autorregulado, por su parte, se divide en uso de estrategias cognitivas y autorregulación.

Duncan y McKeachie (2005), dos protagonistas directos del proceso de creación del MSLQ, llevaron a cabo una revisión que incluyó 56 artículos en los que se empleaba el MSLQ. Estos autores identificaron solo tres artículos en los que se utilizaba la versión abreviada (44 ítems), 22 en los que se empleaba la versión completa (81 ítems) y otros 15 trabajos en los que se usaba o bien el bloque de motivación (7) o bien el bloque de estrategias de aprendizaje autorregulado (8). Estos datos hacen evidente lo señalado por Hilpert, Stempien, Van der Hoeven-Kraft y Husman (2013), que en pocos estudios se han examinado todas las subescalas en una única administración del MSLQ y en menos trabajos se ha estudiado la versión abreviada completa, objeto de la presente validación.

Según ha evidenciado la revisión realizada por Mayer, Faber y Xu (2007), el MSLQ es el

instrumento más utilizado para evaluar la motivación en contextos educativos. Así, el MSLQ se ha convertido en un instrumento estándar en la investigación en el campo de la motivación y la autorregulación (Schunk, 2005).

El cuestionario MSLQ ha sido traducido a diferentes idiomas y validado en múltiples contextos (Erturan-Ilker, Arslan, & Demirhan, 2014; Jakešová & Hrbáčková, 2013; Ramirez-Dorantes, Canto y Rodríguez, Bueno-Alvarez, & Echazarreta-Moreno, 2013; Lee, Yin, & Zhang, 2010; Rao & Sachs, 1999). Además, se ha utilizado como medida criterio en estudios de validez convergente de nuevos instrumentos (este es el caso de Magno, 2011) y de instrumentos ya existentes como el LASSI (Learning and Study Strategies Inventory) en el caso del estudio llevado adelante por Muis, Winne y Jamieson-Noel (2007).

También se ha estudiado la capacidad del MSLQ para predecir el rendimiento académico, siendo el trabajo más relevante al respecto el meta-análisis efectuado por Credé y Phillips (2011). Los autores encontraron que las subescalas de regulación del esfuerzo, autoeficacia y gestión del tiempo y el ambiente de estudio muestran una relación (entre débil y moderada) con el rendimiento. Por su parte, las estrategias más específicas relacionadas con la actividad de estudio (como repetición, elaboración, organización, aprendizaje entre pares, etc.) son las menos predictivas del rendimiento académico. Esto ha sido explicado de distintas maneras por Credé y Phillips (2011).

En primer lugar señalan que algunas estrategias de aprendizaje, como por ejemplo el aprendizaje entre pares, pueden guardar una relación curvilínea con el rendimiento académico que no llega a ser capturada por los coeficientes de correlación. Tanto los estudiantes de bajo como de alto rendimiento pueden no recurrir al aprendizaje entre pares por distintas razones: los primeros pue-

den no estar motivados para hacerlo, en tanto que los segundos pueden considerar no necesitarlo. Otra explicación planteada por estos autores para entender la falta de relación entre las estrategias de aprendizaje específicas y el rendimiento académico tiene que ver con el modo en que se evalúa el rendimiento académico. Frecuentemente, los estudiantes no necesitan resolver pruebas empleando estrategias de aprendizaje que supongan la necesidad de un procesamiento profundo de la información (p. ej.: elaboración). Los autores plantean de modo atinado que las evaluaciones de múltiple opción pueden requerir memorización, pero no evalúan el pensamiento crítico, y los estudiantes que se comprometen con el pensamiento crítico no necesariamente obtienen mejores resultados en las pruebas de múltiple opción.

Pintrich et al. (1993), al cierre del artículo en el cual presentaban la versión completa del MSLQ (81 ítems), señalaron:

Estas significativas aunque moderadas relaciones con las notas de los cursos, son razonables, dados todos los otros factores que se relacionan con las notas de los cursos a nivel universitario y que no son medidos por el MSLQ (las notas individuales del curso en sí mismas no son medidas confiables del rendimiento o del aprendizaje) (p. 812).

Estas consideraciones muestran que, más allá de la complejidad de las relaciones entre los constructos del MSLQ, se sitúa la problemática de la evaluación del rendimiento en el nivel académico.

Al retomar el tema del volumen de producción asociado al MSLQ, se constata otro conjunto de estudios en los que se han analizado las propiedades psicométricas de este instrumento (Pintrich et al., 1993; Hilpert et al., 2013; Dunn, Lo, Mulvenon, & Sutcliffe, 2012); de modo sistemático este instrumento ha mostrado ser una me-

dida confiable de la motivación y las estrategias de aprendizaje autorregulado. Como lo señalaran [Duncan y McKeachie \(2005\)](#), el MSLQ se caracteriza por poseer una estructura conceptual sólida que brinda sustento a su construcción.

Sin embargo, también cabe mencionar que varios estudios han informado problemas vinculados a la estructura factorial latente y han sugerido importantes modificaciones. En particular [Hilpert et al. \(2013\)](#), tras reespecificar la estructura factorial del MSLQ, y luego de explorar varios modelos mediante análisis factorial confirmatorio, propusieron remover más de la mitad de las subescalas, de modo que el modelo final quedó compuesto por tres factores con dos subescalas cada uno: *expectativa* (autoeficacia y control del aprendizaje), *valor* (metas intrínsecas y valor de la tarea) y *autorregulación* (regulación metacognitiva y regulación del esfuerzo).

[Dunn et al. \(2012\)](#) se plantearon el objetivo de evaluar críticamente las dos subescalas de autorregulación, a saber: *autorregulación metacognitiva* y *regulación del esfuerzo*. Los resultados de dicho estudio mostraron que la estructura de dos subescalas de autorregulación no tendría suficiente soporte. Los autores plantearon, por lo tanto, dos subescalas modificadas: la escala general de estrategias de aprendizaje (EGEA) y la escala de estrategias de clarificación para el aprendizaje (ECA). Por su parte, [Hilpert et al. \(2013\)](#) señalaron que la combinación propuesta por [Dunn et al. \(2012\)](#) mostró un buen ajuste con sus datos, e indicaron que esta modificación podría resultar útil a otros investigadores interesados en evaluar la autorregulación. Ambos estudios se llevaron a cabo con la versión completa del MSLQ (81 ítems; [Pintrich et al., 1991](#); [Pintrich et al., 1993](#)).

En la presentación de la versión abreviada que fue objeto de la presente revisión, [Pintrich y De Groot \(1990\)](#) no informaron los análisis factoriales realizados, se limitaron a mencionar que:

“El análisis factorial fue utilizado para guiar la construcción de la escala, y resultó en la exclusión de algunos ítems debido a la falta de correlación o de una estructura factorial estable” (p. 34). Los autores señalaron haber comenzado con 56 ítems, y la escala abreviada quedó finalmente conformada por 44 ítems. Si bien presentaron los análisis de confiabilidad, de correlación entre subescalas y las correlaciones del MSLQ con rendimiento académico, no se dispone de los análisis factoriales exploratorios ni confirmatorios de la versión abreviada. Algunos autores han criticado esta falta de especificación del modelo al mismo tiempo que han señalado problemas con la estructura latente ([Dunn et al., 2012](#)).

Más allá de los análisis críticos del MSLQ, el instrumento ha sido utilizado en cientos de estudios y en muy diversos contextos, mostrando sistemáticamente su consistencia como medida de la motivación y las estrategias de aprendizaje autorregulado. También se ha señalado su capacidad de predecir el rendimiento académico ([Hilpert et al., 2013](#)), siendo esto último uno de los principales intereses de Pintrich y su equipo.

A 27 años de su creación, el MSLQ sigue suscitando un interés permanente, lo que muestra su vigencia, por este motivo, se consideró importante contar con una validación al español de la versión abreviada de este instrumento para el contexto uruguayo.

Método

Participantes

La muestra utilizada en esta investigación estuvo integrada por 318 estudiantes de primer año de la Licenciatura en Psicología (Universidad de la República), 253 mujeres (80%) y 65 hombres (20%). La media de edad de los sujetos fue de 23.38 años, con un desvío típico de 4.2 años.

Si bien se trabajó con una muestra de conveniencia, se realizó un muestreo por turnos para alcanzar mayor representatividad. Para esto se aplicó el cuestionario en el salón de clases en las cuatro franjas horarias en las que se brindaba la unidad curricular obligatoria Psicología del Desarrollo (segundo semestre del plan de estudios de la Licenciatura en Psicología, 2013). La muestra final mantiene las proporciones de la distribución por sexo y la media de edad que caracterizan a los estudiantes que ingresan cada año a la Facultad de Psicología.

Instrumentos

El MSLQ en su versión abreviada contiene 44 ítems que se responden con una escala de tipo Likert de 1 a 7. Dicho cuestionario se divide en dos bloques, denominados *motivación* y *estrategias de aprendizaje autorregulado*. A continuación se presentan las subescalas y los coeficientes alfa de Cronbach presentados por Pintrich y De Groot (1990). El bloque de motivación incluye las subescalas: autoeficacia ($\alpha = .89$), valor intrínseco ($\alpha = .87$) y ansiedad ante las pruebas ($\alpha = .75$). Por otro lado, el bloque de estrategias de aprendizaje autorregulado incluye las siguientes subescalas: uso de estrategias cognitivas ($\alpha = .83$) y autorregulación ($\alpha = .74$).

Se administró la versión uruguaya del MSLQ-UY a todos los sujetos participantes. Se mantuvo la estructura de la versión original en inglés integrada por los siguientes 44 ítems, distribuidos en dos bloques (Tablas 1 y 2).

Rendimiento Académico. Como medida de rendimiento académico se tomó la nota final obtenida por los participantes en el curso Psicología del Desarrollo. Dicha nota surgió del promedio de las notas obtenidas por los estudiantes en dos pruebas parciales.

Tabla 1

Distribución de ítems del Bloque Motivación según subescala.

Bloque Motivación	N° de Ítems	Total
Valor Intrínseco	1, 4, 5, 7, 10, 14, 15, 17, 21	9
Autoeficacia	2, 6, 8, 9, 11, 13, 16, 18, 19	9
Ansiedad ante las Pruebas	3, 12, 20, 22	4
Total		22

Tabla 2

Distribución de ítems del Bloque Estrategias de Aprendizaje Autorregulado según subescala.

Bloque Estrategias de Aprendizaje Autorregulado	N° de Ítem	Total
Uso de Estrategias Cognitivas	23, 24, 26, 28, 29, 30, 31, 34, 36, 39, 41, 42, 44	13
Autorregulación	25, 27, 32, 33, 35, 37, 38, 40, 43	9
Total		22

Procedimiento

Traducción al español de la versión abreviada del MSLQ. El proceso de traducción y adaptación transcultural de la versión abreviada del MSLQ fue realizado según lo propuesto por la Comisión Internacional de Tests (International Test Commission [ITC], 2005), por Van Widenfelt, Treffers, de Beurs, Siebelink y Koudijs (2005) y por Muñiz, Elosua y Hambleton (2013). Participaron en total seis traductores competentes, quienes trabajaron por separado. Dos traductoras bilingües, ambas con formación en motivación en contextos educativos, tradujeron el instrumento del idioma inglés al español. Una vez obtenidas ambas versiones otras dos traductoras bilingües, ambas con formación en psicología y educación y hablantes nativas de inglés, realizaron la traducción inversa de las mismas. En este proceso de retrotraducción se identificaron ítems problemáticos que sufrieron

modificaciones en el proceso de vuelta al idioma original. Por lo tanto, se solicitó a otros dos traductores profesionales, quienes no conocían este instrumento de medición, la traducción directa del inglés al español de los ítems problemáticos. La versión resultante del instrumento fue administrada a 40 estudiantes de Psicología, a quienes se les solicitó su opinión sobre el instrumento en general, así como respecto a la redacción de los ítems. Este paso fue fundamental para la realización de pequeños ajustes. A partir de este proceso se obtuvo la versión en español del MSLQ.

Finalmente, el instrumento en versión papel fue administrado a los estudiantes que participaron del estudio. La aplicación se realizó en el contexto natural del salón de clases de Psicología del Desarrollo, unidad curricular obligatoria del primer año de la Licenciatura en Psicología. Los estudiantes completaron el MSLQ-UY en presencia de la docente del curso antes del comienzo de la clase; fueron informados de los objetivos del estudio e invitados a participar de manera libre y voluntaria, y manifestaron su aceptación mediante la firma de un consentimiento libre e informado. Este estudio contó con el aval del Comité de Ética en Investigación de la Facultad de Psicología. La aplicación del cuestionario insumió entre 20 y 25 minutos.

Análisis Psicométricos. Los datos fueron ingresados y procesados utilizando el paquete estadístico SPSS versión 15.0 (IBM Corporation, 2006) y en primer lugar se analizaron datos descriptivos de la población (sexo y edad). Para llevar adelante el estudio psicométrico de la escala se tomaron como referencia las indicaciones de Carretero-Dios y Pérez (2005). Se realizó un análisis de los ítems para verificar si se ajustaban adecuadamente a la escala.

Se verificó el promedio de los ítems y su desviación estándar con respecto a la media de la

escala. De acuerdo con las recomendaciones de Carretero-Dios y Pérez (2005) se consideraron adecuados aquellos ítems con una desviación típica superior a 1 y una media situada alrededor del punto medio de la escala (simetría próxima a 0).

Para calcular el nivel de discriminación de cada ítem se recurrió al coeficiente de correlación corregido entre la puntuación del ítem y la total obtenida en la dimensión a la que este pertenece. Se consideraron adecuados valores entre 0.25-0.30 (Nunnally & Bernstein, 1995).

Se verificaron los coeficientes de asimetría y se consideraron adecuados aquellos valores entre -1 y $+1$ (Ferrando & Anguiano-Carrasco, 2010).

Para analizar la consistencia interna de la escala se obtuvieron los coeficientes alfa de Cronbach para los dos bloques y sus correspondientes dimensiones. Para estudiar la dimensionalidad de la escala se llevó adelante un análisis factorial exploratorio (AFE) mediante la técnica de extracción por componentes principales y rotación varimax. Para confirmar el número de factores a extraer se consideró el gráfico de sedimentación (Cattell, 1966).

Resultados

Estructura factorial

Se calculó la medida de adecuación muestral KMO (Kaiser-Meyer-Olkin) para comprobar si los valores muestrales permitían o no la aplicación del análisis factorial. El estadístico KMO varía entre 0 y 1; valores menores que .5 indican que los datos que se están analizando no son aptos para efectuar un análisis factorial. La medida KMO fue .806, por lo cual se procedió a realizar el análisis.

Por otro lado, la prueba de esfericidad de Bartlett contrasta la hipótesis nula de que la matriz de correlaciones es una matriz de identidad.

Tabla 3
Análisis factorial exploratorio.

Dimensión	Contenido del Ítem	Factor				
		1	2	3	4	5
<i>Valor Intrínseco</i> ($\alpha = .83$)	17. Creo que es interesante lo que estamos aprendiendo en este curso.	.848				
	15. Creo que lo que estoy aprendiendo en este curso es útil que lo sepa.	.822				
	5. Me gusta lo que estoy aprendiendo en este curso.	.763				
	21. Es importante para mí entender lo que se da en este curso.	.738				
	4. Es importante para mí aprender lo que se da en este curso.	.686				
	7. Creo que seré capaz de usar lo que estoy aprendiendo en este curso, en otros.	.595				
	1. Prefiero las actividades del curso que son desafiantes así puedo aprender cosas nuevas.	.385				
	43. Me esfuerzo para obtener una buena nota, incluso cuando un curso no me gusta.		.725			
<i>Autorregulación Metacognitiva y del Esfuerzo</i> ($\alpha = .64$)	23. Cuando estudio para un parcial o examen, trato de integrar la información de la clase con la de los libros (u otros materiales).		.508			
	33. Incluso cuando los materiales de estudio son aburridos y poco interesantes, continúo trabajando hasta terminar.		.530			
	30. Cuando estudio para un parcial o examen trato de recordar la mayor cantidad de cosas que puedo.		.458			
	24. Cuando hago las tareas, trato de recordar lo que el profesor dijo en clase para contestar las preguntas correctamente.		.458			
	44. Cuando leo trato de conectar lo que estoy leyendo con lo que ya sabía.		.452*			
	10. A menudo para mis trabajos elijo temas de los cuales aprenderé algo, aún cuando requieran de una mayor preparación y esfuerzo de mi parte.		.380*			
	31. Cuando estudio copio mis apuntes de nuevo porque me ayuda a recordar lo que estoy estudiando.		.340*	.709*		
<i>Estrategias Generales de Aprendizaje</i> ($\alpha = .71$)	41. Cuando leo los materiales para este curso, digo las palabras una y otra vez para mí mismo/a, para ayudarme a recordarlos.					
	34. Cuando estudio para un parcial o examen, practico diciendo las cosas importantes para mí mismo/a una y otra vez.			.674		
	35. Antes de comenzar a estudiar, pienso en las cosas que necesitaré hacer para aprender.			.642		
	28. Cuando estudio digo las ideas importantes con mis propias palabras.			.568		
	36. Para hacer nuevas tareas asignadas utilizo lo que he aprendido de tareas anteriores y de la bibliografía del curso.			.433*		
	40. Cuando estoy leyendo, paro cada tanto y vuelvo sobre lo que he leído.			.426		
	25. Me hago preguntas a mí mismo/a para asegurarme que sé sobre lo que he estado estudiando.			.397		
	39. Cuando estoy estudiando para un tema, intento integrar todas las ideas del mismo.			.388		
	42. Hago esquemas de los materiales para ayudarme a estudiar.			.314		

Dimensión	Contenido del Ítem	Factor				
		1	2	3	4	5
Ansiedad y preocupación ante la dificultad ($\alpha = .67$)	26. Es difícil para mí decidir cuáles son las ideas más importantes de los textos que leo.				.729	
	3. Me pongo tan nervioso durante un examen o parcial que no puedo recordar los temas que he aprendido.				.663	
	12. Tengo una sensación de incomodidad y malestar cuando hago un parcial y/o examen.				.616	
	27. Cuando el trabajo es difícil, me doy por vencido o sólo realizo las partes fáciles.		-		.499*	
Autoeficacia ($\alpha = .63$)	9. En comparación con otros compañeros de este curso, creo que soy un buen estudiante.					.768
	16. Mis habilidades de estudio son excelentes en comparación con otros estudiantes de este curso.					.747
	18. En comparación con otros compañeros creo que sé mucho sobre los contenidos de este curso.					.702
	2. En comparación con otros compañeros de este curso espero que me vaya bien.					.474
	13. Creo que voy a obtener una buena nota en este curso.					.432
	32. Hago ejercicios y contesto las preguntas que aparecen al final de cada capítulo aún cuando no tengo que hacerlo.					.366

Nota. Método de extracción: análisis de componentes principales; método de rotación: normalización varimax con Kaiser. La rotación ha convergido en 7 iteraciones. * Ítems que presentan solapamiento en dos dimensiones. Se los ubicó en la dimensión con mayor carga factorial y sentido teórico.

Si el nivel crítico es mayor a .05 no es posible rechazar la hipótesis nula de esfericidad; por lo tanto, no es posible asegurar que el modelo factorial sea adecuado para explicar los datos. Se obtuvo un valor de $p < .001$ en el test de esfericidad de Bartlett. La hipótesis nula fue rechazada, por lo que fue procedente aplicar el análisis factorial.

Se efectuó un análisis factorial exploratorio (AFE), con el método de componentes principales y rotación varimax. El gráfico de sedimentación (Cattell, 1966) permitió visualizar cinco factores principales. El punto de corte en la tendencia descendente se tomó como criterio para la determinación del número óptimo de factores que deben estar presentes en la solución factorial (Pardo & Ruiz, 2002).

Los resultados de este estudio brindan evidencia a favor de la estructura de cinco dimensiones de la escala original. La totalidad de los ítems presentaron cargas por encima de .30, y se

observaron valores aceptables en las comunalidades de todos los ítems ($h^2 > .20$), siendo la menor comunalidad .21 (ítem 38) y la mayor comunalidad .69 (ítem 15). Al llevar adelante el primer AFE, la matriz de componentes rotados evidenció varios ítems con cargas compartidas: 6, 8, 11, 14, 19, 27, 32, 37. Por su parte, el ítem 37 presentó el coeficiente de correlación inter-ítem más bajo de su componente ($r = -.036$) y el ítem 22 presentó el más bajo coeficiente de correlación inter-ítem del bloque motivacional ($r = .007$). Se eliminaron los ítems solapados y el ítem 22 por su baja capacidad de discriminación. Tras la eliminación de estos ítems se realizó un nuevo AFE.

Se encontraron para la estructura factorial del MSLQ-UY cinco factores que explican el 44.28% de la varianza:

Factor 1 (valor intrínseco): ítems 1, 4, 5, 7, 15, 17, 21 (siete de los nueve originales), este factor resultó el más estable de la estructura MSLQ.

Factor 2 (autorregulación –metacognitiva y del esfuerzo–): ítems 10, 23, 24, 29, 30, 31, 33, 43, 44.

Factor 3 (estrategias generales de aprendizaje): ítems 25, 28, 34, 35, 36, 39, 40, 41, 42.

Factor 4 (ansiedad y preocupación ante la dificultad): ítem 3, 12, 20, 26, 27.

Factor 5 (autoeficacia): ítems 2, 9, 13, 16, 18, 32.

Consistencia interna

Se procedió a calcular la confiabilidad del MSLQ-UY por componente y por subescala a partir de la estructura factorial resultante. Los resultados se presentan a continuación.

Tabla 4

Análisis de confiabilidad por subescala según componentes del MSLQ-UY.

	Coefficiente α
Componente Motivacional	.76
Valor Intrínseco	.83
Autoeficacia	.63
Ansiedad ante las Pruebas	.67
Componente Aprendizaje Autorregulado	.75
Estrategias Generales de Aprendizaje	.71
Autorregulación	.64

Los índices de confiabilidad de los componentes fueron aceptables para su uso con fines de investigación ($\alpha = .76$ y $\alpha = .75$).

Correlaciones entre factores

Se estudiaron las relaciones entre los componentes motivacionales y de aprendizaje autorregulado del MSLQ-UY, y los resultados pueden observarse en la Tabla 5. Altos niveles de autoe-

ficacia y valor intrínseco se relacionaron con altos niveles de uso de estrategias cognitivas y de autorregulación. El uso de estrategias cognitivas se relacionó con la autorregulación, y esta fue la correlación más elevada entre dimensiones del MSLQ-UY.

Tabla 5

Correlaciones entre dimensiones motivacionales y de aprendizaje autorregulado.

Dimensión	1	2	3	4
1. Valor Intrínseco	-			
2. Autoeficacia	.443**	-		
3. Ansiedad ante las Pruebas	.036	-.038	-	
4. Uso de Estrategias	.402**	.246**	.129*	-
5. Autorregulación	.322**	.375**	-.057	.545**

Nota. * $p < .05$ (bilateral), ** $p < .01$ (bilateral).

Validez predictiva

A los efectos de estudiar la validez predictiva del MSLQ-UY se tomó como criterio externo el rendimiento académico de los estudiantes (operativizado como la nota final promedio obtenida en el curso donde el instrumento fue aplicado). Se encontró una correlación positiva y significativa entre la subescala de autorregulación y la nota obtenida por los estudiantes ($r = .154$; $p < .01$). El resultado es similar al obtenido mediante un modelo de análisis de regresión múltiple. La única variable de MSLQ-UY que incidió significativamente en el rendimiento, pero de manera muy moderada, fue la subescala de autorregulación ($R^2 = .037$, $\beta = .17$; $p < .01$). Aquellos estudiantes que informaron mayores niveles de autorregulación del aprendizaje mostraron en este estudio una tendencia a obtener un mejor rendimiento.

Validez convergente

Se estudió la validez convergente del MSLQ-UY con respecto a la versión española (Núñez, Martín-Albo, & Navarro, 2005) de la Échelle de Motivation en Éducation (EME; Vallerand, Blais, Brière, & Pelletier, 1989). La EME-E consiste en 28 ítems distribuidos en 7 subescalas de cuatro ítems cada una, que evalúan los tres tipos de motivación intrínseca (relativa al conocimiento, al logro y a las experiencias estimulantes), los tres tipos de motivación extrínseca (externa, introyectada e identificada) y la amotivación (Ryan & Deci, 2000).

Se aplicó la EME-E a una submuestra de 119 participantes que habían completado el MSLQ-UY, 94 mujeres (79%) y 25 varones (21%). La media de edad de este grupo fue de 22.6, con una desviación típica de 6.8 años.

Se calcularon las correlaciones entre las subescalas del MSLQ y las subescalas que integran la EME-E. También se efectuaron análisis de regresión múltiple tomando como variable dependiente cada una de las escalas de la EME-E. Los resultados se presentan a continuación.

Subescala de Motivación Intrínseca al Conocimiento (MIC) de EME-E. El análisis de regresión muestra que esta subescala recibió una incidencia significativa de grado moderado-bajo ($R^2 = .21$) de la subescala de autorregulación de MSLQ-UY ($\beta = .22, p < .001$), la que a su vez correlacionó con MIC ($r = .40, p \leq .01$).

Subescala de Motivación Intrínseca al Logro (MIL) de EME-E. El análisis de regresión mostró una capacidad predictiva moderada-baja ($R^2 = .18$) de las subescalas de autorregulación ($\beta = .26; p < .023$) y ansiedad ante las pruebas ($\beta = .20; p < .023$) de MSLQ-UY sobre esta subescala. Por su parte, MIL correlacionó con autorregulación ($r = .33, p \leq .01$) y con ansiedad ante las pruebas ($r =$

$.20, p \leq .05$).

Subescala de Motivación Intrínseca a las Experiencias Estimulantes (MIEE). Volvió a aparecer un modelo de regresión con capacidad predictiva moderada-baja ($R^2 = .16$) en el que sólo había una variable predictora, la subescala de autorregulación del MSLQ-UY ($\beta = .30; p < .01$). Los resultados mostraron una correlación de autorregulación con MIEE ($r = .38, p \leq .01$).

Subescala de Regulación Externa de EME-E. Se verificó un modelo de regresión que explicaba una parte moderada-baja de la varianza ($R^2 = .15$) con dos variables predictoras: autorregulación ($\beta = -.41; p < .001$) y autoeficacia ($\beta = .24; p < .021$).

Subescala de Regulación Introyectada de EME-E. Se obtuvo un modelo con capacidad predictiva moderada-baja ($R^2 = .14$) con una variable predictora: ansiedad ante las pruebas ($\beta = .32; p < .001$). Se encontró también correlación entre ambas subescalas ($r = .34, p \leq .01$).

Valor intrínseco del MSLQ-UY. Correlacionó con las tres subescalas de motivación intrínseca (MI) de la EME-E: motivación intrínseca al conocimiento ($r = .34, p \leq .01$), al logro ($r = .20, p \leq .01$) y a las experiencias estimulantes ($r = .25, p \leq .01$).

Uso de estrategias cognitivas del MSLQ-UY. Correlacionó con los tres tipos de motivación intrínseca: al logro ($r = .31, p \leq .01$); al conocimiento ($r = .27, p \leq .01$) y a las experiencias estimulantes ($r = .26, p \leq .01$).

Autorregulación del MSLQ-UY. También correlacionó con los tres tipos de motivación intrínseca: al logro ($r = .33, p \leq .01$); al conocimiento ($r = .40, p \leq .01$) y a las experiencias estimulantes ($r = .38, p \leq .01$).

Regulación Externa. Es la más extrínseca de las motivaciones extrínsecas en el continuo de autodeterminación (Ryan & Deci, 2000), y no co-

rrelacionó con ninguna de las dimensiones del MSLQ-UY.

Amotivación de EME-E. Correlacionó de modo negativo con uso de estrategias cognitivas del MSLQ-UY ($r = -.21, p \leq .05$) y también correlacionó negativamente con autorregulación ($r = -.24, p \leq .05$).

Los resultados muestran la validez convergente entre dos instrumentos ampliamente utilizados para evaluar la motivación en contextos educativos.

Discusión

El objetivo de este trabajo fue presentar el proceso de traducción al español, adaptación y validación del MSLQ para el contexto universitario uruguayo. En este trabajo se presenta la primera etapa del trabajo llevado adelante, el cual se ha caracterizado por un riguroso proceso de traducción y adaptación del instrumento desde su versión original. Con respecto a la versión uruguaya del MSLQ (MSLQ-UY), tanto para el bloque *motivación* ($\alpha = .76$) como el bloque *estrategias de aprendizaje autorregulado* ($\alpha = .75$) se obtuvieron índices de consistencia interna que hacen recomendable su uso con fines de investigación. La menor confiabilidad del bloque *motivación* con respecto al instrumento original podría deberse a la disminución de la consistencia interna de la subescala de autoeficacia ($\alpha = .63$) con respecto a la original ($\alpha = .89$) informada por Pintrich y De Groot (1990).

La estructura factorial de cinco dimensiones propuesta por Pintrich y De Groot (1990) fue confirmada en este estudio; sin embargo fue necesario eliminar un total de 7 ítems y la versión final quedó conformada por 37 ítems. Las dos subescalas del bloque *estrategias de aprendizaje autorregulado* fueron las que presentaron más

dificultades con respecto a su estructura factorial, ya que se encontraron varios ítems con cargas repartidas entre ambos componentes. La subescala de autorregulación ($\alpha = .64$) fue categorizada en este estudio como *autorregulación del esfuerzo* y *autorregulación metacognitiva*, dado que contiene ítems que evalúan distintas facetas del aprendizaje autorregulado, y en la versión completa del MSLQ (Pintrich et al., 1993) integra distintas dimensiones que evalúan la regulación del esfuerzo por un lado y la autorregulación metacognitiva por otro. La necesidad de reorganización de las subescalas del bloque EAA ha sido señalada en varios estudios previos (Rao & Sachs, 1999; Lee et al., 2010; Dunn et al., 2012; Hilpert et al., 2013).

Se estudió la relación entre los componentes motivacionales y de aprendizaje autorregulado; las relaciones encontradas van en la misma dirección que las reportadas por Pintrich y De Groot (1990). Los estudiantes que se consideran más capaces (niveles elevados de autoeficacia) presentan mayores niveles de valor intrínseco y es más probable que utilicen estrategias cognitivas y de autorregulación ante la actividad de estudio. Sin embargo, en nuestro estudio no se encontraron las relaciones negativas entre ansiedad ante las pruebas y autoeficacia halladas por Pintrich y De Groot (1990). Esta discrepancia fue previamente informada por Rao y Sachs (1999) en uno de los pocos estudios que propuso validar la versión abreviada del MSLQ, en ese caso para el contexto chino. Estos autores señalan diferencias culturales entre los estudiantes asiáticos y los norteamericanos: en estudiantes chinos, la ansiedad ante las pruebas podría no afectar las creencias sobre sí mismos como aprendices del mismo modo en que afecta a los estudiantes norteamericanos.

Por su parte, Pintrich y De Groot (1990) no informaron relaciones entre ansiedad ante las pruebas y uso de estrategias cognitivas, mientras

que en nuestro estudio ambas dimensiones mostraron estar positivamente relacionadas. Esta discrepancia podría deberse a diferencias culturales entre la muestra uruguaya y la muestra norteamericana; estudiantes ansiosos ante las evaluaciones podrían acudir en mayor medida al uso de estrategias cognitivas para la preparación de las evaluaciones. La relación positiva entre ansiedad ante las evaluaciones y uso de estrategias cognitivas también fue informada en el estudio con población china ya mencionado. Rao y Sachs (1999) plantearon en dicho trabajo que la ansiedad ante las evaluaciones podría tener cierto componente facilitador que lleve al uso eficiente de estrategias cognitivas.

Asimismo, la dimensión de uso de estrategias cognitivas contiene ítems que apuntan al uso de la repetición como estrategia de memorización (p. ej., Ítem 41, *Cuando leo los materiales de este curso, digo las palabras para mí mismo una y otra vez para ayudarme a recordarlos*). Estudios previos con el MSLQ han mostrado diferencias culturales en el empleo de la memorización a nivel universitario. Lee et al. (2010) plantearon que para los estudiantes chinos, la memorización y la comprensión se encuentran culturalmente relacionadas, en tanto que la cultura occidental por el contrario, tiende a disociar memorización de comprensión. Purdie y Hattie (1996) compararon el uso de estrategias de aprendizaje autorregulado en estudiantes japoneses y australianos, y encontraron que los estudiantes japoneses tendían a utilizar la memorización en mayor grado que los estudiantes australianos. Sin embargo, los estudiantes japoneses consideraban la memorización como un camino hacia la comprensión. Los educadores occidentales asumen que la comprensión es una estrategia superior a la repetición, y tienden a disociar ambos procesos; sin embargo las investigaciones con estudiantes de Asia Oriental han mostrado que la memorización para esos

aprendices difiere cualitativamente de la memorización mecánica, y los estudiantes son alentados a memorizar para alcanzar un mayor comprensión (Rao & Sachs, 1999).

Se estudió la validez predictiva del MSLQ-UY tomando como criterio externo el rendimiento académico de los estudiantes en el curso donde el instrumento fue aplicado. Aquellos estudiantes que informaron mayores niveles de autorregulación del aprendizaje tendieron a presentar un mejor rendimiento académico. Estos resultados son consistentes con los reportados por Pintrich y De Groot (1990). La incidencia de esta relación fue moderada-baja, similar a las relaciones de este tipo informadas en estudios previos con MSLQ (Pintrich & De Groot, 1990; Pintrich et al., 1993; Rotgans & Schmidt, 2010; Credé & Phillips, 2011).

Para entender las débiles o moderadas relaciones entre las dimensiones del MSLQ y el rendimiento académico es importante recordar, en primer lugar, lo planteado por Pintrich et al. (1993), quienes puntualizaron que dichos resultados son razonables a la luz de todas las variables que inciden en el rendimiento y que no son evaluadas por dicho instrumento. Al mismo tiempo, los creadores del MSLQ alertaron acerca de que las calificaciones obtenidas en los cursos pueden no ser, por sí mismas, medidas confiables del rendimiento o el aprendizaje.

Como se señaló anteriormente, Credé y Phillips (2011) plantearon como hipótesis que las estrategias de aprendizaje más específicas (como el aprendizaje entre pares) podrían tener una relación curvilínea con el rendimiento que los coeficientes de correlación no logran capturar. Los autores también subrayaron el hecho de que las evaluaciones con formato de opción múltiple —que se aplican sobre todo al inicio de las carreras universitarias— evalúan de modo muy superficial el aprendizaje de los contenidos y no requerirían

del uso de estrategias de procesamiento profundo de la información.

Al estudiar la validez convergente del MSLQ-UY con la Escala de Motivación Educativa se encontró que las relaciones entre las subescalas de ambos instrumentos van en la dirección teórica esperada. *Valor intrínseco* del MSLQ-UY correlacionó con las tres subescalas de motivación intrínseca de la EME-E: motivación intrínseca al conocimiento, al logro y a las experiencias estimulantes, lo cual tiene sentido dada la cercanía conceptual de dichos constructos. Pintrich y De Groot (1990) mencionan que la dimensión *valor* involucra las metas de los estudiantes al abordar una tarea y las creencias acerca de la importancia y el interés de dicha tarea; reconocen que esta dimensión puede ser conceptualizada de diferentes modos, como metas de aprendizaje versus metas de rendimiento, orientación intrínseca versus extrínseca, valor de la tarea o interés intrínseco. Los ítems del MSLQ fueron adaptados, según los autores, desde diversos instrumentos, entre los que mencionan la Escala de Orientación Intrínseca vs. Extrínseca en el Salón de Clases de Harter (1981).

Por su parte, *uso de estrategias cognitivas* del MSLQ-UY correlacionó con los tres tipos de motivación intrínseca (logro, conocimiento y experiencias estimulantes). La subescala *autorregulación* del MSLQ-UY también correlacionó con los tres tipos de motivación intrínseca. Estos resultados se relacionan con los obtenidos por Pintrich y De Groot (1990), y Pintrich et al. (1993), los cuales indican que altos niveles de valor intrínseco se relacionan con el uso de estrategias cognitivas y la autorregulación. Otros estudios han señalado que los estudiantes motivados intrínsecamente tienden a utilizar estrategias de procesamiento más profundo de la información (Karagüven, 2012). Maralani (2016), en un estudio en el cual se utilizaron ambos instrumentos, halló un rol mediador de la motivación intrínseca

en la autorregulación metacognitiva; según este estudio, los ambientes educativos promotores de creatividad incrementarían la motivación intrínseca y, en consecuencia, el uso de estrategias de autorregulación metacognitivas por parte de los estudiantes.

En el presente estudio, *ansiedad ante las pruebas* del MSLQ-UY correlacionó con la subescala de motivación extrínseca *regulación introyectada*. En este tipo de orientación motivacional el estudiante actúa por la presión proveniente del exterior, y realiza las actividades educativas para evitar sentirse culpable y como modo de preservar su autoestima (Ryan & Deci, 2000). A nivel teórico tiene sentido que la ansiedad ante las pruebas prediga este estilo de regulación, que es en parte externo.

Regulación externa (la más extrínseca de las motivaciones extrínsecas en el continuo de autodeterminación) no correlacionó con ninguna de las dimensiones del MSLQ-UY, lo cual tiene sentido si se considera que este instrumento en su versión abreviada no contiene la dimensión orientación extrínseca, limitación que había sido señalada por Lee et al. (2010).

Por último, la subescala *amotivación* de EME-E correlacionó de modo negativo con *uso de estrategias cognitivas* de MSLQ-UY y también correlacionó negativamente con *autorregulación*. La amotivación es definida como el estado de falta de intención para actuar (Ryan & Deci, 2000), de modo que estas correlaciones van en la dirección esperada. Como Pintrich y De Groot (1990) señalaron, los estudiantes no sólo tienen que disponer de ciertas estrategias de aprendizaje, sino que también necesitan estar motivados para emplearlas.

En síntesis, se ha presentado evidencia de la validez convergente entre dos instrumentos ampliamente utilizados para evaluar la motivación en contextos educativos. Especialmente se

encontró que la subescala de autorregulación del MSLQ-UY presenta claras relaciones con los factores de motivación intrínseca del EME-E, y ansiedad ante las pruebas con los factores más extrínsecos de dicha escala.

Conclusiones

Se verificaron índices de consistencia interna para la versión uruguaya del MSLQ que hacen recomendable su uso con fines de investigación, si bien aún no se ha comprobado la conveniencia de su uso con fines diagnósticos. Se presentaron evidencias de su validez estructural y de su validez convergente con una medida motivacional ampliamente utilizada (EME-E).

Es posible afirmar que el MSLQ-UY es un instrumento adecuado para evaluar la motivación académica y las estrategias de aprendizaje autorregulado que emplean los estudiantes universitarios. No obstante, para permitir su uso más allá de los contextos de investigación y aportar a la evaluación y la intervención psicoeducativa se requiere continuar trabajando en la mejora de este instrumento.

En este trabajo se identificaron problemas con la subescala de *autoeficacia*, la cual requiere mejoras en la adaptación cultural de aquellos ítems que apuntan a evaluar este constructo en función de un criterio competitivo que es habitual en el sistema educativo norteamericano, pero que resulta poco ajustado al perfil de los estudiantes de una universidad pública latinoamericana de libre acceso. Por ejemplo, el ítem 9 propone: *En comparación con otros compañeros de este curso creo que soy un buen estudiante*. Se recomienda, para futuras investigaciones en nuestro contexto, la reformulación de estos ítems orientada a la evaluación del componente centrada en el propio desempeño.

Entre las limitaciones de este estudio se encuentra el haber trabajado con estudiantes de una única carrera; es necesario continuar analizando las propiedades psicométricas del MSLQ-UY, ampliando la muestra a estudiantes de otras carreras universitarias e incluyendo asignaturas que utilicen distintos formatos de evaluación. Asimismo, se deberán efectuar otros análisis, en particular un análisis factorial confirmatorio para poner a prueba la estructura teórica del instrumento.

En lo que refiere a la medición del aprendizaje autorregulado, el MSLQ se ha convertido en una medida estándar; sin embargo enfrenta, como todo cuestionario de autoinforme, el problema de ser administrado antes o después de una tarea y no durante la realización de la misma. Esto conlleva dificultades asociadas a la falta de conciencia de los procesos autorregulatorios, a la reconstrucción a partir de la memoria para dar una respuesta sobre una tarea pasada, problemas de comprensión lectora, etc. (Veenman, 2011). Estudios futuros deberían poner en relación el MSLQ con otros tipos de medidas del aprendizaje autorregulado (observacionales, entrevistas, protocolos de pensamiento en voz alta, etc.).

Por último, es necesario mejorar la adecuación del MSLQ al modo en que los estudiantes abordan la actividad de estudio en un contexto donde la tecnología ocupa un lugar cada vez más relevante en la vida cotidiana, ya que ninguna subescala de este instrumento permite evaluar el aprendizaje mediado por nuevas tecnologías (Curione & Huertas, 2016). Los cambios sociales ocurridos desde la aparición del MSLQ justifican la necesidad de este tipo de revisión (Dunn et al., 2012).

Más allá de los ajustes que es necesario realizar, el MSLQ se presenta como una medida confiable que ha mostrado una vez más su posibilidad de adaptación a distintos contextos educativos.

Referencias

- Carretero-Dios, H., & Pérez, C. (2005). Normas para el desarrollo y revisión de estudios instrumentales. *International Journal of Clinical and Health Psychology, 5*(3), 521-551. Recuperado de <http://www.aepc.es/ijchp>
- Credé, M., & Phillips, L. A. (2011). A meta-analytic review of the Motivated Strategies for Learning Questionnaire. *Learning and Individual Differences, 21*(4), 337-346. doi: 10.1016/j.lindif.2011.03.002
- Cattell, R. B. (1966). The scree test of the number of significant factors. *Multivariate Behavioral Research, 1*, 140-161. Recuperado de http://www.tandfonline.com/doi/abs/10.1207/s15327906mbr0102_10
- Curione, K., & Huertas, J. A. (2016). Revisión del MSLQ: Veinticinco años de evaluación motivacional. *Revista de Psicología, 12*(24), 55-67. Recuperado de www.uca.edu.ar/revistadepsicologia
- Duncan, T. G., & McKeachie, W. J. (2005). The making of the Motivated Strategies for Learning Questionnaire. *Educational Psychologist, 40*(2), 117-128. doi: 10.1207/s15326985ep4002_6
- Dunn, K. E., Lo, W.-J., Mulvenon, S. W., & Sutcliffe, R. (2012). Revisiting the Motivated Strategies for Learning Questionnaire: A theoretical and statistical reevaluation of the metacognitive self-regulation and effort regulation subscales. *Educational & Psychological Measurement, 72*(2), 312-331. doi: 10.1177/0013164411413461
- Erturan-Ilker, G., Arslan, Y., & Demirhan, G. (2014). A validity and reliability study of the Motivated Strategies for Learning Questionnaire. *Educational Sciences: Theory Practice, 14*(3), 829-833. doi: 10.12738/estp.2014.3.1871
- Ferrando, P. J., & Anguiano-Carrasco, C. (2010). El análisis factorial como técnica de investigación en psicología. *Papeles del Psicólogo, 31*(1), 18-33. Recuperado de <http://www.papelesdelpsicologo.es/>
- Harter, S. (1981). A new self-report scale of intrinsic versus extrinsic orientation in the classroom: Motivational and informational components. *Developmental Psychology, 17*(3), 300-312. doi: 10.1037/0012-1649.17.3.300
- Hilpert, J. C., Stempien, J., Van der Hoeven-Kraft, K. J., & Husman, J. (2013). Evidence for the latent factor structure of the MSLQ: A new conceptualization of an established questionnaire. *SAGE Open, 3*(4), 1-10. doi: 10.1177/2158244013510305
- IBM. (2006). *SPSS* (Versión 15) [software]. Armonk, NY: IBM Corp.
- International Test Commission (2005). International Guidelines on Test Adaptation. Recuperado de <http://www.intestcom.org>
- Jakešová, J., & Hrbáčková, K. (2014). The Czech adaptation of Motivated Strategies for Learning Questionnaire (MSLQ). *Asian Social Science, 10*(12), 72-78. doi: 10.5539/ass.v10n12p72
- Karagüven, M. H. Ü. (2012). The adaptation of Academic Motivation Scale to Turkish. *Educational Sciences: Theory & Practice, 12*(4), 2611-2618. Recuperado de <http://files.eric.ed.gov/fulltext/EJ1002866.pdf>
- Lee, J. C.-K., Yin, H. B., & Zhang, Z. H. (2010). Adaptation and analysis of Motivated Strategies for Learning Questionnaire in the Chinese setting. *International Journal of Testing, 10*(2), 149-165. doi: 10.1080/15305050903534670
- Magno, C. (2011). Validating the Academic Self-regulated Learning Scale with the Motivated Strategies for Learning Questionnaire (MSLQ) and Learning and Study Strategies Inventory (LASSI). *The International Journal of Educational and Psychological Assessment, 7*(2), 56-73. Recuperado de <https://sites.google.com/site/tijepa2012/home>
- Maralani, F. M. (2016). The mediation role of intrinsic and extrinsic motivation in the relationship between creative educational environment and metacognitive self-regulation. *Journal of Education and Learning, 5*(3), 272-277. doi: 10.5539/jel.v5n3p272
- Mayer, J. D., Faber, M. A., & Xu, X. Y. (2007). Seventy-five years of motivation measures (1930-2005): A descriptive analysis. *Motivation and Emotion, 31*(2),

- 83-103. doi: 10.1007/s11031-007-9060-2
- Muis, K. R., Winne, P. H., & Jamieson-Noel, D. (2007). Using a multitrait-multimethod analysis to examine conceptual similarities of three self-regulated learning inventories. *British Journal of Educational Psychology*, 77(1), 177-195. doi: 10.1348/000709905X90876
- Muñiz, J., Elosua, P., & Hambleton, R. K. (2013). Directrices para la traducción y adaptación de los tests (2ª ed.). *Psicothema*, 25(2), 151-157. Recuperado de <http://www.psicothema.com>
- Nunnally, J. C., & Bernstein, I. (1995). *Teoría psicométrica*. Madrid, España: McGraw-Hill.
- Núñez J. L., Martín-Albo, J., & Navarro, J. G. (2005). Validación de la versión española de la Échelle de Motivation en Éducation. *Psicothema*, 17(2), 344-349. Recuperado de <http://www.psicothema.com>
- Pardo, A., & Ruiz, M. (2002). *SPSS 11 Guía para el análisis de datos*. Madrid: Mc Graw Hill.
- Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. En M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of Self-regulation* (pp. 451-502). San Diego, CA: Academic Press. doi: 10.1016/B978-012109890-2/50043-3
- Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self regulated learning components of classroom academic performance. *Journal of Educational Psychology*, 82(1), 33-40. doi: 10.1037/0022-0663.82.1.33
- Pintrich, P. R., Smith, D. A., García, T., & McKeachie, W. (1991). *A Manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ)*. Ann Arbor: University of Michigan. Recuperado de <https://eric.ed.gov/?id=ED338122>
- Pintrich, P. R., Smith, D. A. F., García, T., & McKeachie, W. J. (1993). Reliability and predictive validity of the Motivated Strategies for Learning Questionnaire (MSLQ). *Educational and Psychological Measurement*, 53(3), 801-813. doi: 10.1177/0013164493053003024
- Purdie, N., & Hattie, J. (1996). Cultural differences in the use of strategies for self-regulated learning. *American Educational Research Journal*, 33(4), 845-871. Recuperado de http://www.jstor.org/stable/1163418?seq=1#page_scan_tab_contents
- Ramirez-Dorantes, M. C., Canto y Rodríguez, J. E., Bueno-Alvarez, J. A., & Echazarreta-Moreno, A. (2013). Psychometric validation of the Motivated Strategies for Learning Questionnaire, with Mexican university students. *Electronic Journal of Research in Educational Psychology*, 11(1), 193-214. Recuperado de <http://www.investigacion-psicopedagogica.org/revista/new/index.php>
- Rao, N., & Sachs, J. (1999). Confirmatory factor analysis of the Chinese version of the Motivated Strategies for Learning Questionnaire. *Educational & Psychological Measurement*, 59(6), 1016-1029. doi: 10.1177/00131649921970206
- Rotgans, J. I., & Schmidt, H. G. (2010). The Motivated Strategies for Learning Questionnaire: A measure for students' general motivational beliefs and learning strategies? *The Asia-Pacific Education Researcher*, 19(2), 357-369. doi: 10.3860/taper.v19i2.1603
- Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, 25(1), 54-67. doi: 10.1006/ceps.1999.1020
- Schunk, D. H. (2005). Self-regulated learning: The educational legacy of Paul R. Pintrich. *Educational Psychologist*, 40(2), 85-94. doi: 10.1207/s15326985ep4002_3
- Vallerand, R. J., Blais, M. R., Brière, N. M., & Pelletier, L. G. (1989). Construction et validation de l'Échelle de Motivation en Éducation (EME) [Construction and validation of the Motivation toward Education Scale]. *Canadian Journal of Behavioural Science/Revue Canadienne des Sciences du Comportement*, 21(3), 323-349. doi: 10.1037/h0079855
- Van Widenfelt, B. M., Treffers, P. D. A., De Beurs, E., Siebelink, B. M., & Koudijs, E. (2005). Translation and cross-cultural adaptation of assessment instruments used in psychological research with

children and families. *Clinical Child and Family Psychology Review*, 8(2), 135-147. doi: 10.1007/s10567-005-4752-1

Veenman, M. V. J. (2011). Alternative assessment of strategy use with self-report instruments: A discussion. *Metacognition and Learning*, 6(2), 205-211. doi: 10.1007/s11409-011-9080-x

Wolters, C. A. (2003). Regulation of motivation: Evaluating an underemphasized aspect of self-regulated learning. *Educational Psychologist*, 38(4), 189-205. doi: 10.1207/S15326985EP3804_1

Funcionamiento de la Categoría Central en Ítems de Confianza para la Matemática

The Role of a Central Category among Items in a Confidence in Mathematics Scale

Facundo Juan Pablo Abal * ^{1,2}, Sofía Esmeralda Auné ^{1,2}, Gabriela Susana Lozzia ²,
Horacio Félix Attorresi ²

1 - Consejo Nacional de Investigaciones Científicas y Técnicas, Argentina.

2 - Universidad de Buenos Aires, Argentina.

Introducción
Método
Resultados
Discusión
Referencias

Recibido: 28/04/2017 Revisado: 30/05/2017 Aceptado: 01/06/2017

Resumen

Se comparan las propiedades psicométricas observadas en un test de Confianza para la Matemática al utilizar formatos de respuesta Likert con y sin categoría central. El instrumento mide un conjunto de creencias del estudiante sobre sus dificultades para responder a las habilidades que demanda la matemática. En el estudio participaron 939 estudiantes de psicología (81% mujeres), los cuales completaron el instrumento con 2 formatos Likert de: a) 5 opciones con categoría intermedia *Ni de acuerdo ni en desacuerdo* y b) 6 opciones con dos categorías centrales (*Más bien en desacuerdo* y *Más bien de acuerdo*). La variación de la escala Likert no afectó sustancialmente las evidencias de validez basadas en la estructura interna (análisis factorial confirmatorio y ajuste al Modelo de Crédito Parcial) ni la relación con otras variables. La función de eficiencia relativa reveló que se obtiene similar información para todos los niveles del rasgo.

Palabras clave: *escala Likert, categoría central, Confianza para la Matemática, estudiantes universitarios*

Abstract

This study compares the psychometric properties of a Confidence for Mathematics Scale using Likert response formats with and without a central category. The scale measures a set of student beliefs about their difficulties to respond to the skills required in the field of Mathematics. The participants were 939 Psychology students (81% female), who responded to the items using two response formats: a) a 5-point scale with the central category *Neither agree nor disagree* and b) a 6-point scale with 2 central categories: *Slightly Disagree* and *Slightly Agree*. Variations on the Likert scale did not substantially affect the evidence of validity based on internal structure (confirmatory factor analysis and Partial Credit Model fit) or in relations with other variables. The relative efficiency function revealed that similar information is obtained for all levels of the trait.

Keywords: *Likert scale, central category, Confidence in Mathematics, college students*

*Correspondencia a: Facundo Juan Pablo Abal. Dirección postal: Zuviría 5691 (1439), Ciudad Autónoma de Buenos Aires. Teléfono: 54 11 46014331.

fabal@psi.uba.ar

Cómo citar este artículo: Abal, F. J. P., Auné, S. E., Lozzia, G. S., & Attorresi, H. F. (2017). Funcionamiento de la categoría central en ítems de Confianza para la Matemática. *Revista Evaluar*, 17(2), 18-31. Recuperado de <https://revistas.unc.edu.ar/index.php/revaluar>

Financiación

Esta investigación fue financiada con subsidios de la Universidad de Buenos Aires (UBACyT 2014-17, Código 20020130100320BA) y de la Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT PICT 2011 N° 0826).

Introducción

Funcionamiento de la categoría central en ítems de Confianza para la Matemática

El debate acerca de la conveniencia de incluir o no una categoría central en las escalas Likert ha generado una multiplicidad de opiniones y abordajes metodológicos (Bisquerra & Pérez-Escoda, 2015; Joshi, Kale, Chandel, & Pal, 2015; Tsang, 2012). La recomendación habitual es que la decisión quede supeditada al juicio prudencial del elaborador de la prueba (Morales, 2006). No obstante, en la gran mayoría de los casos se suele escoger un formato Likert con categoría central (principalmente con cinco opciones) justificado más en la tradición y el pragmatismo que en fundamentos psicométricos (Muñiz, García-Cueto, & Lozano, 2005).

Las posturas a favor de su inclusión sostienen que los evaluados tienden a sentirse más a gusto cuando no se ven compelidos a adoptar una decisión frente a temas para los que se encuentran genuinamente indecisos. Además, la incorporación de categorías más sutiles permitiría discriminar mejor sus respuestas, con el aumento consecuente de la detección de diferencias individuales. Esto último ha sido comprobado especialmente con la incorporación de una categoría intermedia en ítems dicotómicos (Morales, 2006). Los detractores de una respuesta intermedia consideran que su inclusión perjudica las propiedades psicométricas del instrumento, dado que su elección podría estar influenciada por la deseabilidad social (Johns, 2005), ciertas características del enunciado del ítem como ambigüedad y descontextualización (Kulas & Stachowski, 2013) y rasgos de personalidad (Murray, Booth, & Molenaar, 2016).

Uno de los principales argumentos en contra de la categoría intermedia cuestiona la presunción de que los evaluados interpretan de forma inva-

riable los cuantificadores lingüísticos que generalmente acompañan a esta categoría de respuesta (Hernández, Drasgow, & González-Romá, 2004). Idealmente, los evaluados responden a la categoría intermedia cuando su postura legítima está entre las opciones de desacuerdo y acuerdo en el continuo de la escala de respuesta. De esta manera sus respuestas reflejarían niveles moderados del constructo. Sin embargo, Kulas y Stachowski (2009) mostraron que la categoría central también es elegida por los evaluados cuando estos no están seguros de sus opiniones, pretenden ocultarlas o no comprenden cabalmente el enunciado del ítem.

En los últimos años, los modelos politómicos de la Teoría de Respuesta al Ítem (TRI) han tomado protagonismo en el análisis del efecto del formato Likert sobre la validez y confiabilidad de los tests, a través de estudios empíricos y con datos simulados (p. ej. Lee & Paek, 2014; Maydeu-Olivares, Kramp, García-Forero, Gallardo-Pujol, & Coffman, 2009). Más específicamente, los estudios se centraron en analizar los ítems con modelos politómicos que, durante el proceso de estimación, no condicionan un orden para los parámetros correspondientes a las categorías de respuesta de los ítems. Tal es el caso de los modelos de respuesta nominal (Thissen & Cai, 2016) o de crédito parcial (Masters, 2016). De esta manera, en los mencionados estudios se intenta corroborar si efectivamente la categoría intermedia resulta eficaz para captar valores centrales del rasgo (Andrich, de Jong, & Sheridan, 1997; González & Espejo, 2003; Hernández, Espejo, & González, 2006; Hernández, Espejo, González, & Gómez, 2001; Murray et al., 2016; Preston, Reise, Cai, & Hays, 2011).

En el presente trabajo se propone analizar el efecto que produce la inclusión de una categoría central en la escala Likert sobre las propiedades psicométricas de un test. Por tratarse de un estudio empírico, se plantea la comparación de

los resultados obtenidos en una prueba que mide la *confianza para la Matemática* en estudiantes de psicología (Abal, Auné, & Attorresi, 2014). Este constructo permite examinar un conjunto de creencias del estudiante sobre sus posibilidades y dificultades para responder a las habilidades requeridas en la actividad matemática. Resulta particularmente relevante en el contexto de carreras universitarias de corte humanístico-social en las cuales los estudiantes tienden a desconfiar de sus capacidades y perjudican así su implicación en el proceso de aprendizaje (Blanco, 2004; Narro, 1997). Las creencias vinculadas a la propia capacidad percibida por el alumno para enfrentar situaciones de manera exitosa presentan una fuerte carga afectiva y juegan un papel esencial en la concreción de logros (Bermejo, 1996; Gil, Blanco, & Guerrero, 2006; Gómez-Chacón, 2005; Gómez-Chacón, Op't Eynde, & De Corte, 2006; McLeod & McLeod, 2002). En estudios sobre las actitudes vinculadas a la matemática se ha incluido sistemáticamente un componente destinado a la medición de la confianza para la matemática recibiendo estas denominaciones tales como Actitud hacia el éxito en matemática (Fennema & Sherman, 1976), Autopercepción (Adelson & McCoach, 2011), Habilidad (Bazán & Sotero, 1998), Autoconcepto matemático (Palacios, Arias, & Arias, 2014) y Confianza (Auzmendi, 1992; Tapia & Marsh, 2004).

Con base en estas consideraciones teóricas y metodológicas y con el fin de analizar el efecto de la inclusión de una categoría central, nuestro objetivo fue comparar las propiedades psicométricas verificadas en la Escala de Confianza para la Matemática (Abal et al., 2014) empleando ítems con formatos de respuesta de cinco y seis opciones.

A modo de objetivos específicos se planteó considerar: a) las variaciones en los valores adoptados en la estimación de parámetros del Modelo

de Crédito Parcial (MCP) de la TRI y en el ajuste de los datos a este modelo, b) las evidencias de validez basadas en la relación con otras variables externas y c) los estudios de confiabilidad clásicos y con TRI.

Método

Participantes

Los participantes que colaboraron en este estudio fueron 939 estudiantes de la carrera de Psicología (81% mujeres y 19% varones) residentes en el área metropolitana de Buenos Aires, Argentina. La edad osciló entre 18 y 59 años, con un valor promedio de 22.2 años ($DE = 5.4$). El 25% debió recurrir al menos una vez la materia Matemática –correspondiente al primer año de la carrera– luego de haber obtenido un aplazo. La calificación promedio de los participantes en esta materia fue de 6.75 ($DE = 1.84$).

Instrumentos

Cuestionario de variables sociodemográficas y académicas. Este cuestionario recaba información acerca de características tales como género, edad y lugar de residencia. Asimismo, indaga sobre las condiciones de aprobación de Matemática: calificación obtenida y cantidad de veces que debió recurrir la materia antes de poder aprobarla.

Escala de Confianza para la Matemática. Esta prueba se compone de ocho ítems redactados en sentido inverso (Figura 1). El formato de respuesta original presenta una escala Likert con seis categorías: *Totalmente en desacuerdo*, *En desacuerdo*, *Más bien en desacuerdo*, *Más bien de acuerdo*, *De acuerdo* y *Totalmente de acuerdo*. Los alumnos con niveles elevados de confianza tienden a percibirse seguros y eficaces ante situaciones en

las que deben aplicar la Matemática. En cambio, los estudiantes con niveles bajos de confianza se reconocen con dificultades para incorporar, recuperar y aplicar lo aprendido. Estudios previos han aportado evidencias sobre validez basadas en la estructura interna: se verificó la unidimensionalidad del constructo mediante un análisis factorial exploratorio y se ajustó satisfactoriamente el Modelo de Respuesta Graduada de la TRI. En cuanto a la confiabilidad, se obtuvo un coeficiente alfa de .90 y un coeficiente de confiabilidad marginal de la TRI de .91 (Abal et al., 2014).

Escala de Afecto hacia la Matemática. Este instrumento mide el interés del estudiante en involucrarse en actividades vinculadas a la Matemática y los sentimientos asociados al uso de sus términos, símbolos y conceptos. La escala cuenta con evidencias de validez factorial obtenidas a partir de un análisis exploratorio y sus ítems han mostrado un ajuste aceptable al Modelo de Crédito Parcial. El análisis de la consistencia interna arrojó un índice alfa de Cronbach de .91 (Abal, Auné, Lozzia, & Attorresi, 2015).

Escala de Utilidad de la Matemática. Esta prueba evalúa un conjunto de creencias con respecto a la aplicabilidad de las nociones matemáticas en la formación académica y el futuro desempeño profesional. La escala fue validada para una población de estudiantes de Psicología y presenta adecuadas propiedades psicométricas. Todos los ítems se ajustaron al Modelo de Respuesta Graduada y la consistencia interna de la prueba medida mediante alfa de Cronbach fue de .89 (Abal, Galibert, Aguerri, & Attorresi, 2014).

Procedimiento de recolección de datos

Se adoptó un procedimiento que se utiliza habitualmente en estudios empíricos con objetivos similares (p. ej. González & Espejo, 2003;

Maydeu-Olivares et al., 2009). Los participantes respondieron un protocolo que contenía los ítems de confianza con dos formatos de respuesta: el formato original con seis categorías, y un formato de cinco opciones que resumía las dos opciones centrales con el anclaje *Ni de acuerdo ni en desacuerdo*. Otros autores han considerado la administración de un único formato de respuesta y un posterior reagrupamiento de las categorías en la etapa de análisis de datos (p. ej. Matell & Jacoby, 1971; Nunes et al., 2008). Sin embargo, consideramos que la estrategia implementada en este estudio permite analizar la reacción de los individuos frente a los anclajes lingüísticos usados en las escalas Likert.

Los estudiantes completaron el protocolo autoadministrado en grupos de aproximadamente 30 personas, coordinados por miembros del equipo de investigación. Durante la aplicación se aleatorizó el orden en que cada individuo respondió a los formatos, y se intercalaron los demás instrumentos administrados a fin de reducir el impacto de la memorización de las respuestas. La participación fue voluntaria, y se dieron garantías del anonimato y la confidencialidad de las respuestas. También se aclaró que la colaboración no tendría consecuencias negativas ni positivas con respecto al rendimiento académico, y que podían cesar de responder en cualquier punto de la evaluación.

Análisis de datos

Unidimensionalidad. A fin de verificar el supuesto de unidimensionalidad requerido por el MCP se implementó un análisis factorial confirmatorio (AFC) con el programa Mplus (Muthén & Muthén, 2010). Respetando el carácter ordinal de los datos, los parámetros fueron estimados con el método robusto de mínimos cuadrados ponderados (*Weighted Least Squares Mean and Variance*

Adjusted, WLSMV) sobre la base de las matrices de correlaciones policóricas. Se analizó el ajuste al modelo considerando los índices de ajuste comparativo (CFI) y de Tucker-Lewis (TLI), la raíz del promedio de las correlaciones residuales al cuadrado (RMSR) y el error medio cuadrático de aproximación (RMSEA). Siguiendo la recomendación de [Abad, Olea, Ponsoda y García \(2011\)](#) se aceptó el ajuste del modelo para valores de $CFI \geq .95$, $TLI \geq .96$, $RMSR \leq .08$ y $RMSEA \leq .08$.

Estimación y Ajuste del MCP. El MCP (Modelo de Crédito Parcial) define la probabilidad que tiene una persona con un nivel de rasgo θ de escoger la categoría h ($h = 0, \dots, m$) en un ítem i con $m + 1$ categorías. Para ello, se define el parámetro de umbral β_{ih} , que describe la cantidad de θ que demanda el ítem para que el evaluado tenga más chances de elegir la categoría h en lugar de la categoría anterior ($h - 1$). Si bien el MCP no condiciona la estimación de los β_{ih} de un mismo ítem para que presenten valores ordenados, es esperable que estos parámetros registren una secuencia creciente que refleje el incremento de las escalas Likert ([Embretson & Reise, 2000](#)).

Formalmente, la probabilidad de elegir h en el ítem i para un individuo con nivel de rasgo θ sigue la función:

$$P_i(h|\theta) = \frac{e^{\sum_{k=0}^h \theta - \beta_{ik}}}{\sum_{j=0}^m e^{\sum_{n=0}^j \theta - \beta_{in}}}$$

para $h = 0, \dots, m$;

Donde se define: $\sum_{n=0}^0 (\theta - \beta_{in}) = 0$

La aplicación del MCP se realizó con el programa Winsteps ([Linacre, 2006](#)). Mediante el método de máxima verosimilitud conjunta se estimaron 32 parámetros de umbral para el formato de cinco opciones, y 40 parámetros para el for-

mato de seis opciones. Así también se estimaron dos valores de θ para cada sujeto, extraídos sobre la base de sus patrones de respuestas a la prueba con los distintos formatos Likert. El ajuste global del modelo se estudió a partir de los estadísticos ajuste próximo (*Infit*) y lejano (*Outfit*), los cuales se interpretan como medias cuadráticas de los residuales no estandarizados (MNSQ). Estos indicadores adoptan un valor de 1 para un ajuste perfecto, y se contempla una zona de ajuste aceptable entre 0.5 y 1.5 tanto para el *Infit* como para el *Outfit* de los sujetos y de los ítems ([Linacre, 2012](#)).

Asociación con otras variables. Se analizó la asociación de diferentes criterios externos con los θ estimados para cada individuo a partir de sus patrones de respuestas a los ítems de cinco categorías. Para criterios externos medidos cuantitativamente (edad, calificación en Matemática, afecto hacia la Matemática y utilidad de la Matemática) se utilizaron coeficientes r de Pearson. En los casos en que el criterio externo era dicotómico (varones vs. mujeres y cursantes vs. recursantes de Matemática) se calcularon los coeficientes d de Cohen (y sus respectivos intervalos de confianza del 95%) para analizar el tamaño del efecto de la diferencia de medias. Idénticos análisis se usaron para estudiar la relación de los mismos criterios externos con los θ estimados a partir las respuestas a los ítems con seis categorías. Posteriormente se procedió a contrastar los resultados obtenidos para cada criterio externo a fin de detectar diferencias estadísticamente significativas en la forma en que se relacionan con ambas estimaciones de θ . Se aplicaron pruebas de diferencia de r de Pearson para muestras relacionadas ([Steiger, 1980](#)) con el fin de comparar los coeficientes de correlación obtenidos en el estudio de la asociación de cada criterio cuantitativo con los distintos formatos Likert. Los coeficientes d de Cohen fueron comparados considerando sus intervalos

Tabla 1.
Supuesto de unidimensionalidad y análisis de consistencia interna.

	Escala Likert	
	5 categorías	6 categorías
Análisis Factorial Confirmatorio		
CFI	.978	.975
TLI	.970	.965
RMSEA [IC 99%]	.075 [.051 - .099]	.072 [.048 - .096]
RMSR	.044	.041
Consistencia interna		
Alfa de Cronbach [IC 99%]	.93 [.92, .94]	.92 [.91, .93]
<i>Greatest Lower Bound</i>	.95	.94

Nota. CFI = Índice de ajuste comparativo, TLI = Índice de Tucker-Lewis, RMSEA = Error medio cuadrático de aproximación, RMSR = Raíz del promedio de los residuos al cuadrado.

de confianza (IC). Siguiendo a [Maydeu-Olivares et al. \(2009\)](#), en este trabajo se asume que no hay diferencia significativa si se registra una superposición de los IC de las d de Cohen.

Medidas de precisión. Se estudió la consistencia interna del conjunto de ítems a partir de los coeficientes alfa de Cronbach y *Greatest Lower Bound* (GLB). En el marco de la TRI se examinaron las funciones de información de los tests para cada formato de respuesta, y la función de eficiencia relativa (FER). La eficiencia relativa se define como el *ratio* entre las funciones de información de cada formato, por lo que permite estimar el aporte de información que realiza un formato Likert por sobre otro.

Resultados

Comparación del supuesto de unidimensionalidad

Los índices de ajuste comparativo mostraron un resultado adecuado para la verificación del modelo unidimensional usando ambos formatos

de respuesta (TLI y CFI > .95). Con respecto a los índices de ajuste absoluto, los valores de RMSR fueron inferiores al criterio de .08, en tanto que los RMSEA reflejaron un cierto desajuste. Estos resultados permiten, en términos generales, aceptar razonablemente el cumplimiento del supuesto de unidimensionalidad bajo las condiciones que establecen ambas escalas Likert (Tabla 1).

Comparación de la estimación y el ajuste del MCP

La estimación de los parámetros para las escalas Likert alcanzó el criterio de convergencia empleando una cantidad razonable de iteraciones. En la Tabla 2 se muestran los índices obtenidos para la evaluación global del ajuste del MCP a los datos de los dos formatos con cinco y seis categorías de respuesta. Para ambas escalas Likert, las medias de *Infit* y *Outfit* tanto para los sujetos como para los ítems resultaron próximas a 1, lo que refleja un ajuste aceptable.

Tabla 2
Evaluación del ajuste global al Modelo de Crédito Parcial.

Escala Likert	Ajuste global de sujetos				Ajuste global de ítems			
	θ	SE	Infit MNSQ	Outfit MNSQ	β_i	SE	Infit MNSQ	Outfit MNSQ
5 opciones								
Media	.21	.53	1.00	1.00	.00	.05	1.00	1.00
DE	1.80	.11	.42	.43	.47	.00	.08	0.10
6 opciones								
Media	.23	.43	1.03	1.02	.00	.04	1.01	1.02
DE	1.49	.12	.53	.53	.35	.00	.06	.07

Nota. θ = Nivel de rasgo; SE = Error de estimación; β_i = promedio de los β_m de un ítem; MNSQ = Media cuadrática de los residuales no estandarizados del ajuste interno (*Infit*) y ajuste externo (*Outfit*).

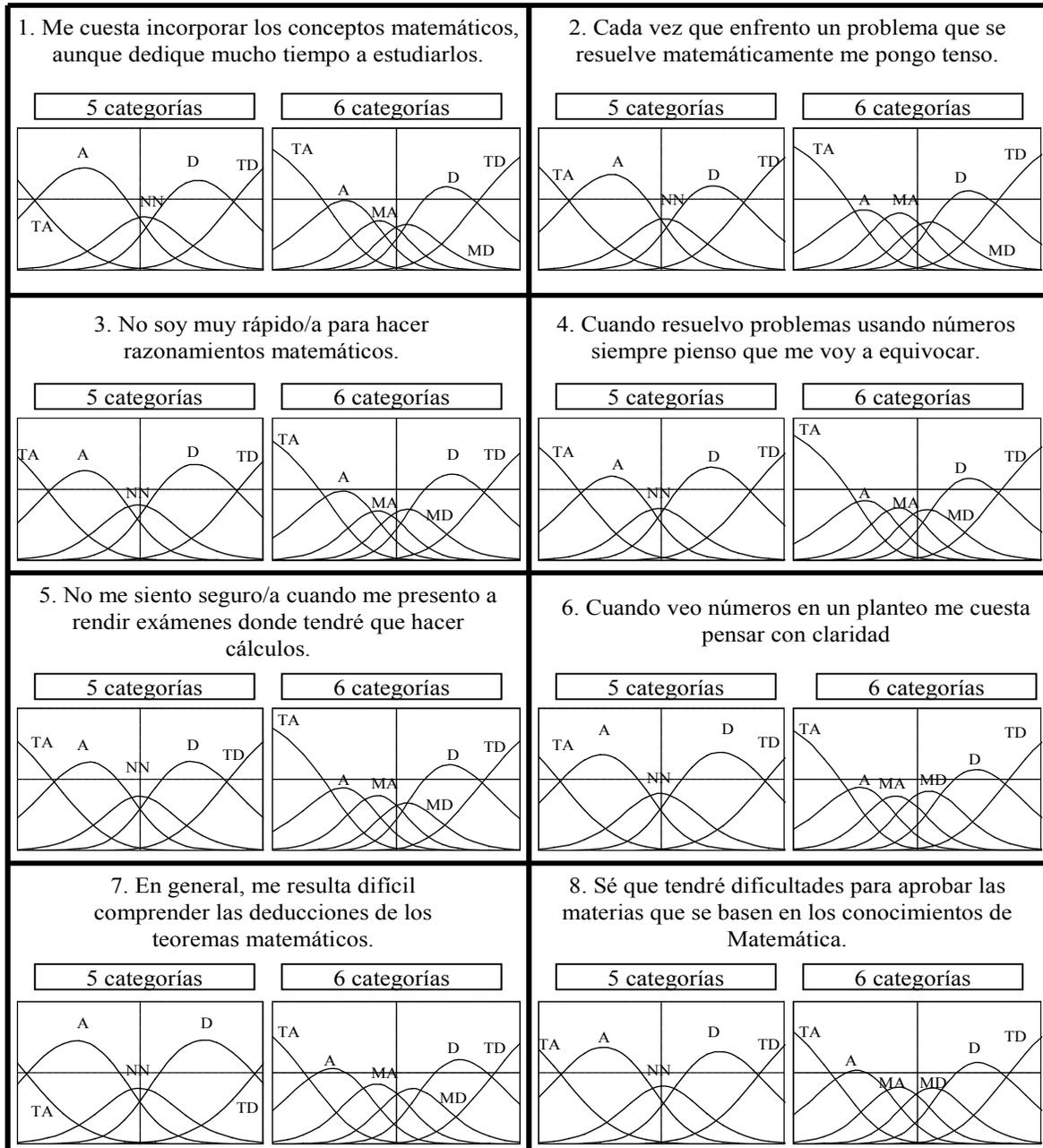
Se encontró una alta correlación entre los θ estimados para cada formato, siendo $r = .91$, $p < .001$ (IC al 99%: .89, .92). Este resultado parece reflejar que el número de opciones podría incidir en la medición del constructo, aunque no sustancialmente. No se encontraron inversiones en los parámetros de umbral β_{ih} de los ítems estimados bajo las dos condiciones de formato. Todas las opciones de las escalas Likert permitieron discriminar algún rango específico del rasgo latente con independencia de la inclusión en la escala de una categoría central (Figura 1). Esto implica que los parámetros de cada ítem presentaron valores crecientes acordes con los niveles de acuerdo reflejados en los anclajes lingüísticos de las escalas Likert.

La categoría central en la escala Likert de cinco opciones mostró una tendencia a ser la más probable para los individuos con niveles de θ intermedios. No obstante, el rango de valores del continuo del rasgo para los que la categoría intermedia mostró más probabilidad que el resto de las categorías de ser elegida resultó bastante acotado. Esto implica que incluso aquellos sujetos ubicados en niveles muy próximos a la media del rasgo también tienen altas probabilidades de optar por

las categorías adyacentes a la central. Como consecuencia, se observó en todos los ítems una clara polarización entre las opiniones de los estudiantes que estaban de acuerdo (categorías *De acuerdo* y *Totalmente de acuerdo*) y las de aquellos que estaban en desacuerdo (categorías *En desacuerdo* y *Totalmente en desacuerdo*). Al comparar estos resultados con las curvas características de los ítems que fueron respondidos con seis opciones se puede apreciar que no se modificó en gran medida la tendencia a elegir la categoría *En desacuerdo* y *Totalmente en desacuerdo*. En cambio, la propensión a estar *De acuerdo* en el formato de cinco categorías parece distribuirse de manera relativamente más equilibrada en las categorías *De acuerdo*, *Más bien de acuerdo* y *Más bien en desacuerdo* en el formato Likert de seis puntos.

Comparación de la asociación con criterios externos

Las asociaciones de los θ con las variables externas resultaron acorde a lo esperado a nivel teórico. Como muestra la Tabla 3, los valores r de Pearson obtenidos al estudiar la correlación de la

**Figura 1**

Curvas características de los ítems respondidos con cinco y seis categorías.

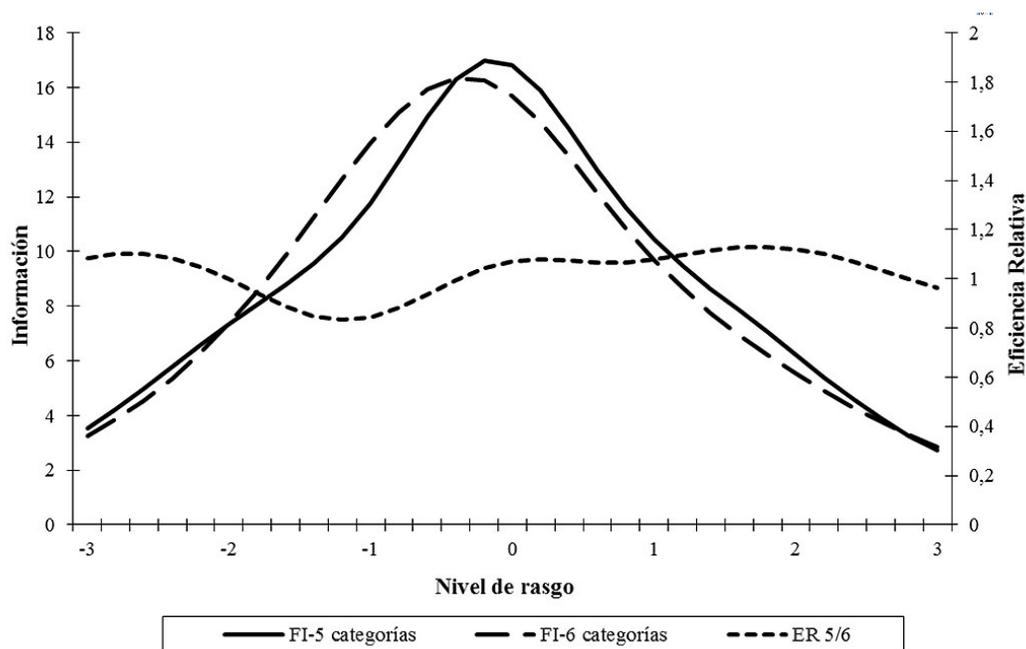
confianza con otras variables (edad, calificación en Matemática, afecto hacia la Matemática y utilidad de la Matemática) no mostraron diferencias estadísticamente significativas entre el formato de respuesta con categoría intermedia y el que no la incluía. Tampoco se encontraron diferencias al

contrastar grupos de datos en base a criterios dicotómicos (varones vs. mujeres, cursantes vs. re-cursantes de Matemática). En efecto, se encontró una superposición entre los IC de los tamaños del efecto d de Cohen de los dos formatos.

Tabla 3

Relación de utilidad con criterios externos.

Criterio externo	Estadístico	Escala Likert		Significación
		5 categorías	6 categorías	
Sexo	d [95% IC]	.37 [.24, .50]	.32 [.20, .41]	---
Edad	r	-.023	-.001	$t_{(936)} = -1.57$, NS.
Recurso Matemática	d [95% IC]	.52 [.39 - .64]	.51 [.39 - .60]	---
Nota en Matemática	r	.266***	.278***	$t_{(936)} = -0.89$, NS.
Afecto hacia la Matemática	r	.675***	.690***	$t_{(936)} = -1.49$, NS.
Utilidad de la Matemática	r	.310***	.326***	$t_{(936)} = -1.21$, NS.

Nota. *** $p < .001$ **Figura 2**

Funciones de información para los formatos de cinco y seis categorías, y función de eficiencia relativa.

Comparación de medidas de precisión

Los índices de consistencia interna alfa de Cronbach y GLB registraron valores altamente satisfactorios y similares para ambos formatos de escala Likert (Tabla 1). Las funciones de información de los tests muestran que el instrumento aporta mayor cantidad de información en los ni-

veles medio y medio-bajo en ambos modelos, de cinco o seis categorías. Este resultado se vio reflejado en la FER, que se mantuvo relativamente constante y próxima a 1 a lo largo de todo el espectro del rasgo latente, lo que indica que la precisión de medición es similar en ambas versiones del test (Figura 2).

Discusión

La especificación del formato de respuesta de los ítems pertenecientes a tests de comportamiento típico supone un conjunto de decisiones que debe tomar quien construye un instrumento. A pesar de esto, es poco frecuente que los autores justifiquen por qué utilizan una cantidad particular de opciones o por qué escogen determinados cuantificadores lingüísticos para acompañar las categorías. Más infrecuente es encontrar que los autores recurran a estudios empíricos para analizar si estas decisiones adoptadas en el diseño de la prueba pueden perjudicar la medición. El objetivo de este trabajo fue estudiar si la inclusión o no de una categoría central en la escala Likert de la Escala de Confianza para la Matemática impacta sobre las propiedades psicométricas del instrumento.

Las diferencias encontradas al analizar las evidencias de validez basadas en la estructura interna no resultaron considerables. Los índices de bondad de ajuste del AFC adoptaron valores óptimos y similares para las mediciones de *confianza* obtenidas a partir de ítems con formato tanto de cinco como de seis categorías. En la misma línea, tampoco se observó una discrepancia en la evaluación del ajuste al MCP. Algunos autores han encontrado que una disminución del número de opciones parece asociarse a una mejora en el ajuste de los modelos de la TRI (Hernández, Muñoz, & García-Cueto, 2000; Lee & Paek, 2014; Maydeu-Olivares et al., 2009). Sin embargo este efecto no se ha percibido en el presente estudio, posiblemente debido a que se torna más notable en la medida en que el ítem presenta menos cantidad de categorías (p. ej. tres o cuatro opciones).

El cambio en la escala Likert tampoco afectó la relación de la *confianza* con otras variables consideradas como criterios externos. Los tamaños de los efectos de las diferencias de medias de

confianza entre varones y mujeres fueron semejantes cuando el constructo se midió usando los ítems con cinco o con seis categorías. Este mismo resultado se corroboró en los tamaños de los efectos en la comparación de los grupos de estudiantes que aprobaron Matemática la primera vez que la cursaron y los que debieron recursarla. Por otra parte, las asociaciones de la *confianza* con las variables edad, calificación en Matemática, afecto hacia la Matemática y utilidad de la Matemática también presentaron resultados similares al cotejar los coeficientes de correlación obtenidos para cada formato de respuesta.

Con respecto a la confiabilidad, la inclusión de la categoría central no produjo un efecto significativo en los índices de consistencia interna calculados en el marco de la teoría clásica. Desde la perspectiva de la TRI, la FER mostró que usando ambos formatos de respuesta se alcanza una precisión similar para todos los niveles del rasgo.

En conclusión, los resultados obtenidos permiten afirmar que el uso de formatos de respuesta con cinco o seis categorías no afecta sustancialmente las evidencias de validez ni confiabilidad de la Escala de Confianza para la Matemática.

El análisis de los ítems de la Escala de Confianza mediante el MCP mostró que la categoría central es eficaz para detectar sujetos que presentan niveles medios del rasgo. En efecto, para todos los ítems se observó un rango de valores dentro del espectro del rasgo en el que las curvas características de la opción *Ni de acuerdo ni en desacuerdo* fueron más probables que las curvas de las demás categorías del ítem. En un estudio previo, empleando el mismo anclaje lingüístico, Rojas y Fernández (2000) encontraron un comportamiento de la curva de la categoría similar al aquí observado en un grupo de universitarios que interpretaban esta opción como una categoría que expresa una posición intermedia. En cambio, un grupo que interpretaba la categoría central en tér-

minos de duda o indecisión mostró parámetros de umbral desordenados y, consecuentemente, curvas de categorías que no resultaban las más probables en ningún intervalo de la escala. La homogeneidad en los resultados obtenidos con el grupo de estudio en los distintos ítems apoya la hipótesis de una tendencia de los sujetos a atribuir un significado acorde a lo esperado para la categoría central de la escala. No obstante, nos parece conveniente complementar este resultado con estudios cualitativos que permitan corroborar los patrones de interpretación de las opciones. Al igual que en los estudios realizados por [Kulas y Stachowski \(2013\)](#), aquí también se podría solicitar a un grupo de los sujetos que informen *en voz alta* el proceso que lleva a la elección de una respuesta al ítem.

Al observar las curvas de los ítems respondidos con la escala Likert de cinco opciones podría aceptarse que la categoría central resulta eficaz en tanto que atrae a una parte no despreciable de sujetos. Sin embargo, para todos los ítems se verifica que la probabilidad de elección de la categoría central disminuye considerablemente al ser comparada con las probabilidades de optar por las categorías adyacentes. Esto implica que, aún teniendo la posibilidad de escoger una respuesta neutra, la mayoría de los alumnos tienden a expresar una posición a favor o en contra de la opinión planteada en el ítem. Posiblemente este resultado se encuentre relacionado con el constructo medido. La confianza del estudiante sobre su destreza matemática, en tanto creencia, se forma sobre la base de su trayectoria académica ([Nuñez et al., 2005](#)). En consecuencia, es esperable que al alcanzar un nivel universitario, el individuo se afirme en sus experiencias previas a fin de percatarse de las expectativas de logro a las que se podría enfrentar, y logre discernir con mayor claridad, al menos en términos dicotómicos, si puede alcanzarlas o no.

Ahora bien, la disminución en la probabili-

dad de elección de la categoría central indica que podría resultar adecuado ensayar un formato de cuatro opciones. En esta línea, en futuros estudios buscaremos analizar si una escala Likert más parsimoniosa, de cuatro categorías, afecta significativamente el grado de ajuste del modelo o la precisión de la medida. Así también, examinaremos si los resultados obtenidos en este trabajo son comparables con los de grupos de universitarios de otras carreras y para la medición de otros constructos.

En cuanto a las limitaciones del presente estudio, conviene señalar que las conclusiones obtenidas sobre el funcionamiento de la categoría central se circunscriben estrictamente a los ítems de la Escala de Confianza para la Matemática. Aún cuando las posibilidades de generalización de estos resultados sean reducidas, los mismos podrían considerarse orientadores respecto a otras escalas que midan un constructo afin o que indaguen en una población con características similares. El diseño metodológico con medidas repetidas implementado en este estudio permite analizar los efectos de variaciones intra-individuales en cuanto a la respuesta de un mismo ítem bajo dos formatos de respuesta diferentes: de cinco y de seis opciones. Esto resulta una ventaja frente a abordajes con datos simulados, los cuales examinan el efecto de la categoría intermedia sobre la validez y la confiabilidad del instrumento, pero sin considerar el proceso psicológico subyacente involucrado en la respuesta a los ítems en un test específico.

Referencias

- Abad, F. J., Olea, J., Ponsoda, V., & García, C. (2011). *Medición en ciencias sociales y de la salud*. Madrid: Síntesis.
- Abal, F., Auné, S., & Attorresi, H. (2014). Comparación del modelo de respuesta graduada y la teoría clásica

- sica de tests en una Escala de Confianza para la Matemática. *Summa Psicológica UST*, 11(2), 101-113. Recuperado de <http://summapsicologica.cl/index.php/summa>
- Abal, F. J. P., Auné, S. E., Lozzia, G. S., & Attorresi, H. F. (2015). Modelización de una prueba de afecto hacia la matemática con la teoría de respuesta al ítem. *Revista de Psicología UCA*, 11(21), 23-34. Recuperado de <http://bibliotecadigital.uca.edu.ar>
- Abal, F. J. P., Galibert, M. S., Aguerri, M. E., & Attorresi, H. F. (2014). Comparación de los modelos respuesta graduada y crédito parcial aplicados a una escala de utilidad de la matemática. *Revista Argentina de Ciencias del Comportamiento*, 6(3), 6-16. Recuperado de <https://revistas.unc.edu.ar/index.php/racc>
- Adelson, J. L., & McCoach, D. B. (2011). Development and psychometric properties of the Math and Me Survey: Measuring third through sixth graders' attitudes towards mathematics. *Measurement and Evaluation in Counseling and Development*, 44(4), 225-247. doi: 10.1177/0748175611418522
- Andrich, D., de Jong, J. H. A. L., & Sheridan, B. E. (1997). Diagnostic opportunities with the Rasch model for ordered response categories. En J. Rost & R. Langeheine (Eds.), *Applications of latent trait and latent class models in the social sciences* (pp. 59-72). Münster, Germany: Waxmann Verlag.
- Auzmendi-Escribano, E. (1992). *Las actitudes hacia la matemática-estadística en las enseñanzas medias y universitarias: Características y medición*. Bilbao, España: Mensajero.
- Bazán, J. L., & Sotero, H. (1998). Una aplicación al estudio de actitudes hacia la matemática en la UNALM. *Anales Científicos UNALM*, 36, 60-72.
- Bermejo, V. (1996). Enseñar a comprender las matemáticas. En J. Beltrán-Llera & C. Genovard-Roselló (Eds.), *Psicología de la Instrucción I* (pp. 256-279). Madrid: Síntesis.
- Bisquerra, R., & Pérez-Escoda, N. (2015). ¿Pueden las escalas Likert aumentar en sensibilidad? *Revista d'Innovació i Recerca en Educació*, 8(2), 129-147. doi: 10.1344/reire2015.8.2828
- Blanco-Blanco, A. (2004). Enseñar y aprender estadística en las titulaciones universitarias de ciencias sociales: Apuntes sobre el problema desde una perspectiva pedagógica. En J. C. Torre-Puente & E. Gil-Coria (Eds.), *Hacia una enseñanza universitaria centrada en el aprendizaje* (pp. 143-190). Madrid: Universidad Pontificia Comillas.
- Embretson, S., & Reise, S. (2000). *Item Response Theory for Psychologists*. Mahwah, NJ: Erlbaum Publishers.
- Fennema, E., & Sherman, J. A. (1976). Fennema-Sherman Mathematics Attitudes Scales: Instruments designed to measure attitudes toward the learning of mathematics by females and males. *Journal for Research in Mathematics Education*, 7(5), 324-326. doi: 10.2307/748467
- Gil-Ignacio, N., Guerrero-Barona, E., & Blanco-Nieto, L. (2006). El dominio afectivo en el aprendizaje de las matemáticas. *Electronic Journal of Research in Educational Psychology*, 4(8), 47-72. Recuperado de <http://investigacion-psicopedagogica.org/revista/new/index.php>
- Gómez-Chacón, I. M. (2005). *Matemática emocional. Los afectos en el aprendizaje matemático*. Madrid, España: Narcea.
- Gómez-Chacón, I. M., Op't Eynde, P., & De Corte, E. (2006). Creencias de los estudiantes de matemáticas. La influencia del contexto de clase. *Enseñanza de las Ciencias*, 24(3), 309-324. Recuperado de <http://ddd.uab.cat/record/18?ln=carecord/14?ln=ca>
- González-Romá, V., & Espejo, B. (2003). Testing the middle response categories "Not sure", "In between" and "?" in polytomous items. *Psicothema*, 15(2), 278-284. Recuperado de <http://www.psicothema.es/>
- Hernández, A., Drasgow, F., & González-Romá, V. (2004). Investigating the functioning of a middle category by means of a mixed-measurement model. *Journal of Applied Psychology*, 89(4), 687-699. doi: 10.1037/0021-9010.89.4.687
- Hernández, A., Espejo, B., & González-Romá, V. (2006).

- The functioning of central categories Middle Level and Sometimes in graded response scales: Does the label matter? *Psicothema*, 18(2), 300-306. Recuperado de <http://www.psicothema.es/>
- Hernández-Baeza, A., Espejo-Tort, B., González-Romá, V., & Gómez-Benito, J. (2001). Escalas de respuesta tipo Likert: ¿Es relevante la alternativa “indiferente”? *Metodología de Encuestas*, 3(2), 135-150. Recuperado de <http://casus.usal.es/pkp/index.php/MdE>
- Hernández, A., Muñiz, J., & García-Cueto, E. (2000). Comportamiento del modelo de respuesta graduada en función del número de categorías de la escala. *Psicothema*, 12(2), 288-291. Recuperado de <http://www.psicothema.es/>
- Johns, R. (2005). One size doesn't fit all: Selecting response scales for attitude items. *Journal of Elections, Public Opinion and Parties*, 15(2), 237-264, doi: 10.1080/13689880500178849
- Joshi, A., Kale, S., Chandel, S., & Pal, D. K. (2015). Likert-scale: Explored and explained. *British Journal of Applied Science & Technology*, 7(4), 396-403. doi: 10.9734/BJAST/2015/14975
- Kulas, J. T., & Stachowski, A. A. (2009). Middle category endorsement in odd-numbered Likert response scales: Associated item characteristics, cognitive demands, and preferred meanings. *Journal of Research in Personality*, 43(3), 489-493. doi: 10.1016/j.jrp.2008.12.005
- Kulas, J. T., & Stachowski, A. A. (2013). Respondent rationale for neither agreeing nor disagreeing: Person and item contributors to middle category endorsement intent on Likert personality indicators. *Journal of Research in Personality*, 47(4), 254-262. doi: 10.1016/j.jrp.2013.01.014
- Lee, J., & Paek, I. (2014). In search of the optimal number of response categories in a rating scale. *Journal of Psychoeducational Assessment*, 32(7), 663-673. doi: 10.1177/0734282914522200
- Linacre, J. M. (2006). *Winsteps® (Version 3.63.0)* [Software]. Beaverton, Oregon: Winsteps.com.
- Linacre, J. M. (2012). *Winsteps® Rasch measurement computer program User's Guide*. Beaverton, Oregon: Winsteps.com.
- Masters, G. N. (2016). Partial credit model. En W. J. van der Linden (Ed.), *Handbook of Item Response Theory, Volume 1: Models* (pp. 109-126). Boca Raton, FL: CRC Press.
- Matell, M. S., & Jacoby, J. (1971). Is there an optimal number of alternatives for Likert scale items? Study I: Reliability and validity. *Educational and Psychological Measurement*, 31, 657-674. doi: 10.1177/001316447103100307.
- Maydeu-Olivares, A., Kramp, U., García-Forero, C., Gallardo-Pujol, D., & Coffman, D. (2009). The effect of varying the number of response alternatives in rating scales: Experimental evidence from intra-individual effects. *Behavior Research Methods*, 41(2), 295-308. doi: 10.3758/BRM.41.2.295
- McLeod, D. B., & McLeod, S. H. (2002). Synthesis - Beliefs and mathematics education: Implications for learning, teaching and research. En G. C. Leder, E. Pehkonen, & G. Törner (Eds.), *Beliefs: A hidden variable in mathematics education?* (pp. 115-126). Dordrecht: Kluwer Academic Publishers.
- Morales, P. M. (2006). *Medición de actitudes en psicología y educación* (3º ed.). Madrid: Universidad Pontificia Comillas.
- Muñiz, J., García-Cueto, E., & Lozano, L. M. (2005). Item format and the psychometric properties of the Eysenck Personality Questionnaire. *Personality and Individual Differences*, 38(1), 61-69. doi: 10.1016/j.paid.2004.03.021
- Murray, A. L., Booth, T., & Molenaar, D. (2016). When middle really means “Top” or “Bottom”: An analysis of the 16PF5 using Bock's nominal response model. *Journal of Personality Assessment*, 98(3), 319-331. doi: 10.1080/00223891.2015.1095197
- Muthén, L. K., & Muthén, B. O. (2010). *Mplus User's Guide* (6º ed.). Los Angeles, CA: Muthén & Muthén.
- Narro-Ramírez, A. E. (1997). Investigación sobre la concepción de la matemática en las ciencias sociales

- en la UAM-Xochimilco. *Política y Cultura*, 9, 249-280. Recuperado de <http://www.redalyc.org/revista.oa?id=267>
- Nunes, C. H. S. S., Primi, R., Farias-Oliveira-Nunes, M., Muniz, M., Freitas-da Cunha, T., & Couto, G. (2008). Teoría de respuesta ao item para optimização de escalas tipo Likert –um exemplo de aplicação. *Revista Iberoamericana de Diagnóstico y Evaluación - e Avaliação Psicológica*, 25(1), 51-79. Recuperado de <http://www.redalyc.org/revista.oa?id=4596>
- Núñez, J. C., González-Pienda, J. A., Alvarez, L., González, P., González-Pumariega, S., Roces, C. ... Rodrigues-Feio, L. do S. (2005). Las actitudes hacia las matemáticas: perspectiva evolutiva. *Actas del VIII Congresso Galaico Português de Psicopedagogia*, 2389-2396. Recuperado de <http://www.educacion.udc.es/grupos/gipdae/index.php?pagina=VIIIcongreso>
- Palacios, A., Arias, V., & Arias, B. (2014). Attitudes towards mathematics: Construction and validation of a measurement instrument. *Revista de Psicodidáctica*, 19(1), 67-91. doi: 10.1387/RevPsicodidact.8961
- Preston, K., Reise, S., Cai, L., & Hays, R. D. (2011). Using the nominal response model to evaluate response category discrimination in the PROMIS emotional distress item pools. *Educational and Psychological Measurement*, 71(3), 523-550. doi: 10.1177/0013164410382250
- Rojas-Tejada, A. J., & Fernández-Prados, J. S. (2000). Análisis de las alternativas de respuestas intermedias mediante el modelo de escalas de clasificación. *Metodología de Encuestas*, 2(2), 171-183. Recuperado de <http://casus.usal.es/pkp/index.php/MdE>
- Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, 87(2), 245-251. doi: 10.1037/0033-2909.87.2.245
- Tapia, M., & Marsh, G. E. (2004). An instrument to measure mathematics attitudes. *Academic Exchange Quarterly*, 8(2), 16-21. Recuperado de <http://www.rapidintellect.com/AEQweb/cho253441.htm>
- Thissen, D., & Cai, L. (2016). Nominal categories models. En W. J. van der Linden (Ed.), *Handbook of Item Response Theory, Volume 1: Models* (pp. 51-73). Boca Raton, FL: CRC Press.
- Tsang, K. K. (2012). The use of midpoint on Likert Scale: The implications for educational research. *Hong Kong Teachers' Centre Journal*, 11, 121-130. Recuperado de <http://www.edb.org.hk/hkct/content.aspx?id=20060614161412&lang=e>

Matrices Progresivas de Raven: Percentiles para Preescolares Cubanos de entre 4 y 6 años

Yaser Ramírez-Benítez * ¹, Bárbara Bermúdez-Monteagudo ², Shuyeng Acea-Vanega ³

1 - Universidad de Cienfuegos, Cuba.

2 - Universidad de Cienfuegos, Cuba.

3 - Sala de Rehabilitación del Policlínico Ramón Claudio Delgado Amistoy, Cienfuegos, Cuba.

Introducción
Método
Resultados
Discusión
Referencias

Recibido: 07/06/2017 **Revisado:** 26/06/2017 **Aceptado:** 14/07/2017

Resumen

La prueba Matrices de Raven es una alternativa para evaluar la inteligencia en niños en edad preescolar. La investigación se propuso determinar percentiles en la prueba Matrices de Raven para niños cubanos de 4 a 6 años, validar percentiles con respecto a un punto de corte cubano (Ramírez-Benítez et al., 2015) y determinar las propiedades psicométricas de la prueba desde un análisis de ítems. Participaron 803 niños (380 varones) seleccionados en 15 instituciones educativas de enseñanza regular en las ciudades de Cienfuegos y Matanzas, Cuba. El análisis de validez entre el punto de corte cubano y los percentiles obtenidos indicó que los percentiles deben utilizarse en la práctica con precaución. No obstante, los percentiles clasifican con mayor flexibilidad al niño con alteraciones en el neurodesarrollo con respecto al punto de corte cubano de Ramírez-Benítez et al. (2015). En resumen, los percentiles son adecuados para clasificar al niño preescolar cubano, aunque requieren de otros estudios de validez.

Palabras clave: *Matrices de Raven, inteligencia fluida, preescolar, neurodesarrollo*

Abstract

The Raven Progressive Matrices Test is an alternative to assess intelligence in preschool children. This study attempted to determine percentiles in the Raven Matrices Test to assess Cuban children between 4 and 6 years old, to validate these percentiles regarding a Cuban cut-off point (Ramírez-Benítez et al., 2015) and to determine the psychometric properties of the test through item analysis. A total of 803 children (380 boys) were selected in 15 regular educational institutions in the cities of Cienfuegos and Matanzas, Cuba. The validity analysis between a Cuban cut-off point and the obtained percentiles indicated that these percentiles should be used in practice with caution. However, percentiles allow for a more flexible assessment in children with neurodevelopmental disorder than the Cuban cut-off point from Ramírez-Benítez et al., 2015. Evidence supports adequacy of percentiles to assess Cuban preschool children, although further validity studies are required to support these findings.

Keywords: *Raven's Matrices Test, fluid intelligence, preschool children, neurodevelopment*

*Correspondencia a: Yaser Ramírez-Benítez. Correo: yaser@citmacfg.cu

Cómo citar este artículo: Ramírez-Benítez, Y., Bermúdez-Monteagudo, B., & Acea-Vanega, S. (2017). Matrices progresivas de Raven: Percentiles para preescolares cubanos de entre 4 y 6 años. *Revista Evaluar*, 17(2), 32-46. Recuperado de <https://revistas.unc.edu.ar/index.php/revaluar>

Introducción

El estudio de la inteligencia es uno de los capítulos más leídos e investigados en la historia de la psicología. El enfoque psicométrico ha sido el modelo teórico que más investigaciones ha generado sobre la inteligencia, y al mismo tiempo ha formulado varias preguntas que aún permanecen sin respuesta.

En la década de los '90, la Asociación Americana de Psicología publicó un informe oficial sobre los conocimientos y el desconocimiento de la inteligencia humana (Neisser et al., 1996), el cual es de obligada consulta para los profesionales de la conducta. Según el comité redactor de ese informe, presidido por Ulric Neisser, la inteligencia se puede medir, y los tests de inteligencia la miden adecuadamente. Dichos tests constituyen el modo de evaluación de la inteligencia más preciso, confiable y adecuado, y tienen múltiples aplicaciones y utilidades.

En dicho informe se reconoce la importancia de la teoría del factor g de Spearman como la teoría de la inteligencia más conocida de todos los tiempos, pues ha posibilitado el diseño de varias pruebas de inteligencia en diversos contextos culturales. De acuerdo con esta teoría, la inteligencia puede ser explicada en función de un factor de capacidad mental general g y de un gran número de factores específicos s (Spearman, 1904). El factor g es una habilidad fundamental que interviene en todas las operaciones mentales, representa la energía mental y se pone en funcionamiento en toda tarea no automatizada. Es una capacidad de reflexión que permite al sujeto observar lo que ocurre en su interior, concebir las relaciones esenciales existentes entre dos o más ideas (educación de relaciones) y captar las ideas implícitas en una relación (educación de correlatos).

Las tareas cognitivas más cargadas de g son aquellas que exigen razonamiento deductivo o

inductivo, visualización espacial, razonamiento cuantitativo, y conocimiento y razonamiento verbal (significado de las palabras, distinciones entre palabras relacionadas, sinónimos-antónimos, analogías verbales y comprensión lectora; Crone et al., 2009). Entre los principales instrumentos que miden el factor g se encuentran las Matrices Progresivas de Raven (Raven, 2004), el Test del Dominó (Anstey, 1963) y los Tests de Factor g de Cattell (Cattell & Cattell, 1994).

El Test de Matrices Progresivas Coloreadas de Raven (TMCR) es un instrumento adecuado para medir el factor g en edades infantiles, y presenta particularmente tres grandes atractivos: (1) es una prueba fácil de aplicar y su ejecución requiere de poco tiempo; (2) puede ser aplicado en poblaciones de niños con alteraciones del neurodesarrollo, en particular en niños con afectaciones del habla (sordera, hipoacusia, afasias, autismo), y (3) las puntuaciones tienen una correlación significativa con el aprendizaje, en especial en los primeros años de vida (Ramírez-Benítez, Torres-Díaz, & Amor-Díaz, 2016).

Desde esta perspectiva, el TMCR tiene una amplia utilidad en contextos clínicos y educativos; sin embargo, desde el año 2000 hasta la fecha, varios grupos de investigación han realizado publicaciones en las que se manifiestan a favor de actualizar las normas de dicha prueba, específicamente para las edades comprendidas entre 6 y 60 años (Bandeira, Boccato-Alves, Giacomel, & Lorenzato, 2004; Fernández-Liporace, Ongarato, Saavedra, & Casullo, 2004; Fernández-Nistal & Mercado-Ibarra, 2014; Ivanovich et al., 2000; Mansilla, Vásquez, & Estrada, 2012; Rossi-Casé et al., 2014, 2015, 2016). Los investigadores justifican la actualización de la norma por tres razones: (1) la necesidad de realizar interpretaciones diagnósticas adecuadas; (2) la necesidad de realizar revisiones periódicas de las pruebas psicométricas en función de los modelos teóricos ac-

tuales o cambios culturales en el contexto, y (3) necesidad de revisar las puntuaciones directas de la prueba debido a la inconsistencia en los resultados que podría ocasionar el efecto Flynn.

En relación con el efecto Flynn, este hace alusión al hecho de que la norma utilizada para evaluar a los sujetos en una población dada se hace obsoleta conforme pasan los años, pues se espera una puntuación mayor a medida que pasa el tiempo. De esta forma se hace necesario realizar investigaciones que permitan modificar las normas y clasificar de manera adecuada la inteligencia.

El aumento de las puntuaciones es mayor en aquellos tests que miden la inteligencia fluida (pruebas de inteligencia libres de influencia educativa) que en las que miden la inteligencia cristalizada (Sundet, Barlaug, & Torjussen, 2004). Esto llevó a pensar que el aumento estaba relacionado con el factor *g* de la inteligencia, ya que las pruebas del factor *g* tienen una mayor correlación con las tareas de inteligencia fluida comparadas con factores específicos como las pruebas de vocabulario, cálculo y lectura.

La causa de este fenómeno se desconoce hasta el momento, pues el aumento de los puntajes directos no es sinónimo de aumento de la inteligencia en la población; además, las evidencias presentadas hasta el momento no son suficientes para afirmar que la inteligencia de la población esté aumentando. Una explicación posible de este fenómeno está relacionada con el entrenamiento vinculado a diversos factores culturales y educativos tales como el incremento de la educación desde edades tempranas y el acceso directo a diversas tecnologías digitales desde edades tempranas (Rossi-Casé et al., 2014).

Una reciente revisión del tema evidencia que el efecto Flynn ocurre en algunas poblaciones de países desarrollados (Flynn & Rossi-Casé, 2012; Rossi-Casé et al., 2015) mientras que otros

resultados no confirman tal efecto, pues destacan que los valores disminuyen con el transcurso de los años o bien que existe un efecto meseta para las puntuaciones obtenidas (Bandeira, Costa, & Arteché, 2012; Flynn, 2013; Rossi-Casé et al., 2014). El principal debate se centra en los países desarrollados, donde se espera un efecto Flynn debido a la influencia del nivel económico, el grado educacional y los niveles de vida. Sin embargo los resultados no son consistentes. En efecto, si en un grupo de países desarrollados no ocurre el efecto Flynn, lo cual constituye la minoría de los casos, se espera que en el resto de los países subdesarrollados no ocurra tal efecto (Teasdale & Owen, 2007; Brouwers, Van de Vijver, & Van Hemert, 2008).

En Cuba se han realizado pocos estudios para proponer una norma del TMCR (Cairo-Valcárcel, Cairo-Martínez, Bouza, & Solazabal, 2000; Ramírez-Benítez, Jiménez-Morales, & Diaz-Bringas, 2015); en su lugar se utilizan normas de la población mexicana (Raven, 2004) por la cercanía cultural a nuestro contexto. En forma similar, no se han realizado estudios sobre el efecto Flynn con muestras cubanas. Por ende, el objetivo principal de este trabajo es obtener percentiles para clasificar la inteligencia de manera adecuada. En este sentido, al considerar la aplicación corta y fácil de la prueba, Cairo-Valcárcel et al. (2000) presentaron una propuesta de percentiles con la intención de valorar la inteligencia en la población cubana, aunque la propuesta se focalizó en jóvenes y adultos. Por otra parte, un intento de establecer un punto de corte para conocer el desarrollo del razonamiento analógico en niños preescolares cubanos fue llevado a cabo por Ramírez-Benítez et al. (2015); sin embargo, el uso de un punto de corte posibilita una clasificación basada en dos opciones: desarrollo típico de la edad versus desarrollo insuficiente para la edad. De esta forma, el profesional está limitado

a clasificar al niño en dos grupos, lo cual es poco habitual en la práctica profesional cuando se trata de un instrumento de inteligencia.

Otros puntos de interés están relacionados con el uso del TMCR en niños en edad preescolar, la actualización de la norma y el análisis de ítems de la prueba para estas edades. Una revisión en la literatura reveló que la mayoría de las actualizaciones de la norma se proponen para la evaluación en sujetos de entre 6 años y edades adultas (Bandeira et al., 2004; Fernández-Liporace et al., 2004; Fernández-Nistal & Mercado-Ibarra, 2014; Ivanovich et al., 2000; Mansilla et al., 2012; Rossi-Casé et al., 2014, 2015, 2016), lo cual resulta curioso si se compara con el aumento reciente de investigaciones con niños en edad preescolar (Tenorio et al., 2012; Lázaro, Ostrosky, & Lozano, 2012; Solovieva, Loredó, Quintanar, & Lázaro, 2013) con el fin de detectar alteraciones del neurodesarrollo, en especial entre los 3 y 6 años de edad. En particular, las investigaciones valoran la inteligencia en preescolares a través de las diferentes variantes de la escala de Wechsler (WPP-SI-IV, Wechsler, 2011, y WISC-IV, Wechsler, 2012), pues esta escala aporta mayor información al valorar el desarrollo cognitivo del niño. Sin embargo, una estimación rápida del razonamiento analógico podría ser realizada a través del TMCR en niños en edad preescolar, ya que sus puntuaciones muestran una significativa relación con el desempeño de las habilidades escolares, tanto al inicio como al final de la etapa escolar.

Si bien el TMCR está diseñado para niños entre 5 y 11 años, algunos investigadores han demostrado que un niño de 4 años y medio de edad es capaz de resolver algunos problemas de matrices (Ramírez-Benítez et al., 2015; Raven, 2004). Asimismo, Raven (2004) relacionó el desarrollo de la capacidad intelectual con la edad cronológica, y estableció que desde los 4 años el niño puede resolver algunos problemas, pues es capaz

de darse cuenta de que a una figura grande le falta una parte y esa parte puede estar dentro de las opciones de respuesta. Así, el TMCR puede ser aplicado desde los 4 años y puede, en consecuencia, estimular el desarrollo del pensamiento, así como detectar alteraciones a edades tempranas.

En relación con el análisis de ítems en la prueba Raven, se puede observar una tendencia similar a la de los estudios de actualización, pues el investigador que realiza un estudio de normalización utiliza también los datos de la muestra para realizar un análisis de ítems. Desde esta perspectiva, se puede observar una carencia de estudios que valoren el rendimiento de los ítems del TMCR en edades preescolares.

Considerando los puntos señalados, este trabajo propone determinar un conjunto de percentiles para la prueba TMCR a fin de clasificar el razonamiento analógico de niños cubanos en edad preescolar, entre 4 y 6 años, así como determinar la validez de los percentiles propuestos con respecto al punto de corte de Ramírez-Benítez et al. (2015). De esta manera, el especialista puede clasificar al niño dentro de un rango de cinco percentiles: percentil 90: niños con puntuaciones superiores para su edad; percentil 75: niños con puntuaciones altas para su edad; percentil 50: niños con puntuaciones típicas para su edad; percentil 25: niños con puntuaciones bajas para su edad, y percentil 10: niños con puntuaciones muy bajas para su edad. Además, este estudio se propone determinar las propiedades psicométricas del TMCR desde un análisis de ítems a fin de valorar tanto la complejidad del instrumento para niños en edad preescolar como la capacidad para discriminar entre niños con baja y alta habilidad en el razonamiento analógico.

Método

Participantes

Se realizó un diseño de tipo no experimental transversal correlacional, en el cual se utilizaron dos muestras de normalización: la muestra de Ramírez-Benítez et al. (2015) y una muestra no probabilística intencionada aplicada en 2016 (de enero a mayo).

En total, participaron 803 niños (380 varones y 423 niñas) de entre 4 años y 5 meses y 6 años y 11 meses de edad seleccionados en 15 instituciones educativas de enseñanza regular en las ciudades de Cienfuegos y Matanzas, Cuba (8 círculos infantiles y 7 escuelas regulares).

En 2016, se seleccionaron al azar 4 círculos infantiles y 3 escuelas regulares de la provincia de Cienfuegos y se evaluó a todos los niños de entre 4 y 6 años. En los círculos infantiles, los niños fueron evaluados por año de vida: 4º año de vida: edades comprendidas entre 4 años y 5 meses y 4 años y 10 meses; 5º año de vida: edades comprendidas entre 5 años y 2 meses y 5 años y 9 meses; y finalmente, 6º año de vida: edades comprendidas entre 6 años y 2 meses y 6 años y 11 meses. Una parte de la muestra se seleccionó en los círculos infantiles debido a que estas son instituciones donde la población infantil recibe orientación educativa de manera sistemática y organizada. De esta forma, se podía controlar la asistencia y permanencia de los niños en la investigación.

En relación con las escuelas regulares, se evaluaron todos los grupos de 1º grado con edades comprendidas entre 6 años y 2 meses y 6 años y 11 meses. Además, en las escuelas regulares la matrícula de niños con 6 años es mayor, lo cual justifica el tamaño mayor de la muestra de niños de 6 años con respecto a las demás edades.

El estudio fue autorizado por el Ministerio de Educación Provincial y se contó con el consentimiento de los padres de aquellos niños evaluados.

Se estima que la muestra utilizada corresponde aproximadamente a un 10% de la población de niños en edad preescolar en Cienfuegos y Matanzas, de acuerdo con el Anuario Estadístico de Salud en el 2013 (Ministerio de Salud Pública, 2013).

Instrumento

Test de Inteligencia de Matrices Progresivas Coloreadas de Raven (Raven, 2004). Esta prueba consiste en responder 36 problemas de tipo no verbal que se presentan en un cuaderno de papel. Cada uno de ellos consta de una figura incompleta que el niño debe completar. Se presentan seis alternativas de respuesta posibles, una de ellas es la respuesta correcta y el resto son distractores. La prueba está dividida en tres series de doce problemas cada una: A, Ab y B. La complejidad se incrementa progresivamente a lo largo de la prueba y a lo interno de las series de modo que A1 es el ítem más fácil y B12 el más difícil, pero A12 es más difícil que B1. Las funciones cognitivas que se examinan principalmente son: percepción de tamaños, percepción de orientación en el espacio, en una dirección y en dos direcciones simultáneamente, aprehensión de figuras discretas especialmente relacionadas con un todo (Series A y Ab), análisis de un todo en sus componentes, capacidad de concebir figuras correlativas y deducción de relaciones (Serie B). Por cada respuesta correcta se otorga un punto y se recoge una puntuación total en la ejecución del test.

Procedimiento

La prueba se aplicó a todos los niños de la muestra de manera individual por la mañana. La aplicación se realizó en un local tranquilo e iluminado en presencia de la educadora.

Análisis de datos

Estadística Descriptiva. Se utilizó el programa SPSS versión 21 (IBM, 2012) para determinar la estadística descriptiva por edades: la media, la desviación estándar, la prueba Kolmogorov-Smirnov de bondad de ajuste a la curva normal y el máximo y el mínimo de las puntuaciones en la prueba.

Propiedades psicométricas. La confiabilidad del instrumento se evaluó a través del análisis de ítems mediante el coeficiente alfa de Cronbach cuando el ítem es eliminado y la correlación ítem-total (índice de homogeneidad). Se utilizó el valor de alfa de Cronbach cuando el ítem es eliminado debido a la importancia que tiene cada ítem de la prueba en niños en edad preescolar, pues se espera que algunos ítems sean más difíciles que otros. Este método analiza el coeficiente alfa de cada ítem, por lo que si alfa aumentara significativamente tras eliminar un ítem específico, esto indicaría que la exclusión de ese ítem aumenta la confiabilidad de la escala. Un ítem con valores de alfa significativamente por encima de los demás podría ser candidato a ser eliminado por ser muy complejo o muy fácil. En igual sentido, se utilizó la correlación ítem-total para conocer cuánto aporta cada ítem al resultado final de la prueba. De esta forma, se obtuvo una medida de confiabilidad de los ítems. La correlación ítem-total se clasificó siguiendo las recomendaciones de Ebel y Frisbie (1986): valores mayores que .39, correlaciones excelentes, el ítem se puede conservar; valores entre .30 - .39, correlaciones buenas, el ítem tiene posibilidades de mejorar; valores entre .20 - .29, correlaciones regulares, el ítem necesita revisión, y valores menores a .20, correlaciones pobres, el ítem se debe descartar o revisar en profundidad.

En igual sentido se obtuvieron los índices de dificultad y de discriminación de los ítems de la

prueba, los cuales fueron clasificados según Crocker y Algina (1986): dificultad baja con valores menores o iguales a .80; dificultad media con valores entre .79 y .50; y dificultad alta con valores menores que .50.

Punto de corte. Se obtuvieron cinco percentiles en cada edad para clasificar al niño con baja y alta habilidad en el razonamiento analógico.

Validez de los percentiles. La validez de los percentiles se evaluó a través de tres procedimientos: (1) la coincidencia (cantidad de verdaderos positivos y negativos) al clasificar el rendimiento intelectual entre el punto de corte de Ramírez-Benítez et al. (2015) y los percentiles propuestos; (2) el cálculo del coeficiente de probabilidad diagnóstica (CPD), y (3) la cantidad de casos clasificados con alteraciones en el rendimiento intelectual por los percentiles propuestos en una muestra de niños con alteraciones en el neurodesarrollo (submuestra de validación, 130 niños, de Ramírez-Benítez et al., 2015). Para aplicar los tres procedimientos fue necesario obtener los indicadores de *sensibilidad* (capacidad del punto de corte para detectar a los casos positivos –niños que presentan alteraciones intelectuales) y *especificidad* (capacidad del punto de corte para descartar a los casos negativos –niños que no presentan alteraciones intelectuales).

El coeficiente de probabilidad diagnóstica estima cuánto más probable es el resultado de una prueba o punto de corte en personas enfermas en relación con las personas sanas (Moratalla-Rodríguez, 2015). El cociente de probabilidad positivo (CPD+), con una significativa relación con los valores de la sensibilidad, estima cuánto más probable es obtener un resultado positivo en sujetos con problemas en el desarrollo que en sujetos sin problemas. Se calcula de la siguiente manera: $(CPD+) = (Se) / (1 - Es)$ donde *Se* representa sensibilidad y *Es* representa especificidad. El cociente de probabilidad negativo (CP-), con una

significativa relación con los valores de la especificidad, estima cuánto más probable es obtener un resultado negativo en sujetos enfermos que en sujetos sanos. Se calcula de la siguiente manera: $(CPD-) = (1 - Se) / (Es)$.

Según Moratalla-Rodríguez (2015) los valores del CPD se clasifican de la siguiente manera:

- CP > 10: incrementos amplios de la probabilidad diagnóstica.
- CP 5 - 10: incrementos moderados de la probabilidad diagnóstica.
- CP 2 - 5: incrementos pequeños de la probabilidad diagnóstica.
- CP 1 - 2: incrementos insignificantes de la probabilidad diagnóstica.
- CP 1: sin cambios de la probabilidad diagnóstica.
- CP 0.5 - 1: descensos insignificantes de la probabilidad diagnóstica.
- CP 0.2 - 0.5: descensos pequeños de la probabilidad diagnóstica.
- CP 0.1 - 0.2: descensos moderados de la probabilidad diagnóstica.
- CP < 0.1: descensos amplios de la probabilidad diagnóstica.

Resultados

Estadística descriptiva

Los valores de la media indicaron que en la medida que incrementaba la edad, los niños aumentan sus puntuaciones en la prueba Raven, tal y como se indica en la Tabla 1.

Con respecto a la desviación estándar, los valores en los niños de 6 años fueron mayores que los de las demás edades. Esto se debió a que 6 niños obtuvieron puntuaciones máximas. Dichos casos fueron detectados a través de un análisis de frecuencia de las puntuaciones obtenidas por edades. Al eliminarse estos 6 casos del análisis se pu-

do observar que el valor de DE disminuyó a 1.85.

La media y la moda en cada edad coinciden como se muestra en la Tabla 1, es por ello que los valores de Z en la prueba de Kolmogorov-Smirnov presentan niveles significativos de normalidad: a los 4 años ($Z = 4.29$, Sig. .33), a los 5 años ($Z = 1.93$, Sig. .20) y a los 6 años ($Z = 2.62$, Sig. .24).

Tabla 1

Estadística descriptiva por edades. Percentiles por edades. Valores de Z para las tres muestras independientes (4 años, 5 años y 6 años) en la prueba de normalidad Kolmogorov-Smirnov.

Variables sociodemográficas	Cantidad de Casos		
	4 años	5 años	6 años
Características de la muestra			
Niños	97	92	191
Niñas	108	113	202
Total	205	205	393
Provincia			
Cienfuegos	107	102	203
Matanzas	98	103	190
Instituciones			
Círculos infantiles	205	95	112
Enseñanza regular	-	110	281
Prueba Raven			
Media	14.01	15.04	16.03
Mediana	14.00	15.00	16.00
Desviación típica	1.44	1.82	2.74
Mínimo	8.00	9.00	6.00
Máximo	17.00	23.00	30.00
Percentiles			
10	10.00	11.00	11.00
25	11.00	12.00	13.00
50	14.00	15.00	16.00
75	15.00	17.00	18.00
90	16.00	18.00	20.00
Z de K-Smirnov	4.29	1.93	2.62
Sig.	.33	.20	.24

Tabla 2
Índices de dificultad y discriminación de los 36 ítems del TMCR.

Índice de Dificultad	Índice de Discriminación		
	Buena (mayor o igual a .20)	Aceptable (entre .10 y .19)	Mala (menor o igual a .10)
Baja (mayor o igual a .80)	8, 9, 10, 13, 18, 24, 25, 34	1, 2, 28, 29, 30	
Media (entre .51 y .79)	4, 5, 6, 7, 19, 20, 21, 22, 23, 27, 33, 35	3, 14, 15, 16, 17	
Alta (menor o igual a .50)	11, 12, 26, 31, 32, 36		

Propiedades psicométricas: Análisis de ítems

Confiabilidad. El coeficiente alfa de Cronbach cuando el ítem es eliminado indicó valores adecuados para la muestra total (entre .84 y .89). El análisis indicó que si se eliminan los ítems 1, 4, y 32 alfa aumenta a .89.

Los valores de alfa también resultaron adecuados por edad. A los 4 años se obtuvieron valores de alfa de entre .82 y .86, con posibilidades de aumentar hasta .86 si se eliminan los ítems 1, 9 y 33. A los 5 años los valores de alfa fueron de entre .85 y .88 y no se encontraron ítems con valores destacados; por ende, no se recomienda eliminar ningún ítem. A los 6 años se obtuvieron valores de entre .84 y .90 con posibilidades de aumentar alfa a .90 si se eliminan los ítems 1, 2, 3, 4 y 32.

En relación con la correlación ítem-total o índice de homogeneidad, las correlaciones oscilaron entre excelentes y buenas con valores de entre .33 y .56: El 30% de los ítems obtuvo excelentes correlaciones (ítems 6, 7, 8, 11, 17, 18, 19, 23, 24, 28, 30) y el resto obtuvo buenas correlaciones (ítems 1, 2, 3, 4, 5, 9, 10, 12, 13, 14, 15, 16, 20, 21, 22, 25, 26, 27, 29, 31, 32, 33, 34, 35, 36).

Índices de dificultad y discriminación. En la Tabla 2 se muestran los índices de dificultad y de discriminación de los ítems de la prueba. En rela-

ción con el índice de discriminación, el 72.2% (26 ítems de 36) de los ítems son buenos para discriminar entre niños con alta y baja habilidad (ítems 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 31, 32, 33, 34, 35 y 36) y un 27.7% (10 ítems de 36) de los ítems son aceptables (ítems 1, 2, 3, 14, 15, 16, 17, 28, 29 y 30).

Por otra parte, el índice de dificultad indicó que el 36% (13 ítems de 36) de los ítems tiene una dificultad baja (ítems 1, 2, 8, 9, 10, 13, 18, 24, 25, 28, 29, 30 y 34), el 47% (17 ítems de 36) de los ítems tiene una dificultad media (ítems 3, 4, 5, 6, 7, 14, 15, 16, 17, 19, 20, 21, 22, 23, 27, 33, 35) y el 16% (6 ítems de 36) dificultad alta (11, 12, 26, 31, 32, 36; véase la Tabla 2).

Validez de los percentiles propuestos. De manera general, la coincidencia al clasificar a los niños entre el punto de corte cubano (Ramírez-Benítez et al., 2015) y los percentiles propuestos es adecuada (69%); no obstante, los valores de la sensibilidad son ligeramente bajos (véase la Tabla 3) debido a la cantidad de verdaderos positivos con respecto a la cantidad de falsos negativos (casos positivos según el punto de corte cubano y negativos por los percentiles propuestos).

A los 4 años la coincidencia resultó ligeramente baja (53%), pues la sensibilidad resultó muy baja con varios falsos negativos (46% de la

Tabla 3

Sensibilidad, especificidad y coeficiente de probabilidad diagnóstica de los percentiles propuestos según el punto de corte cubano.

Edad	Sensibilidad	Especificidad	CPD+	CPD-	VN	VP	FN	FP
4 años	.23	1	.23	.77	79	30	96	0
5 años	.48	1	.48	.52	66	67	72	0
6 años	.50	1	.50	.50	110	144	139	0
Total	.43	1	.43	.57	255	241	307	0

Nota. CDP+ (coeficiente de probabilidad diagnóstica positivo), CPD- (coeficiente de probabilidad diagnóstica negativo), VN (verdaderos negativos), VP (verdaderos positivos), FN (falsos negativos) y FP (falsos positivos).

Tabla 4

Cantidad de niños clasificados según los percentiles propuestos en la submuestra con alteraciones en el neurodesarrollo (130 niños) de Ramírez-Benítez et al., 2015.

Punto de corte cubano	Percentiles propuestos				
4 años	P10 (10)	P25 (11-13)	P50 (14)	P75 (15)	P90 (16)
Positivos (≤ 14)	-	3	-	-	-
Negativos (≥ 15)	-	-	12*	10	5
5 años	P10 (11)	P25 (12-14)	P50 (15-16)	P75 (17)	P90 (18)
Positivos (≤ 16)	-	35	-	-	-
Negativos (≥ 17)	-	-	13*	2	-
6 años	P10 (11-12)	P25 (13-15)	P50 (16-17)	P75 (18-19)	P90 (20)
Positivos (≤ 17)	-	23	-	-	-
Negativos (≥ 18)	-	-	18*	8	1
Total	P10	P25	P50	P75	P90
Positivos	-	61 (46.92%)	-	-	-
Negativos	-	-	43* (33.07%)	20 (15.38%)	6 (4.61%)

Nota. Se han señalado con (*) los falsos negativos según el punto de corte de Ramírez-Benítez et al. (2015). Positivos: niños clasificados con alteraciones en el desarrollo. Negativos: niños clasificados sin alteraciones en el desarrollo según los percentiles propuestos.

muestra; véase la Tabla 3). A los 5 años la coincidencia mostró valores adecuados (65%), aunque la sensibilidad se mantuvo baja debido a la cantidad de falsos negativos (34% de la muestra). A los 6 años la coincidencia resultó adecuada (64%) y en igual sentido la sensibilidad resultó apropiada tal y como se muestra en la Tabla 3, aunque con varios falsos negativos (35% de la muestra).

En relación con el coeficiente de probabilidad diagnóstica, los resultados indicaron que los

percentiles propuestos tienen un descenso pequeño de la probabilidad diagnóstica, ya sea en el CPD+ como en el CPD-, tal y como se muestra en la Tabla 3. Este resultado ocurre cuando se realiza el análisis en la muestra total y por edades.

De manera general, el 46.92% de la submuestra con alteraciones en el neurodesarrollo fue clasificado por el percentil 25, tal y como se muestra en la Tabla 4, el 33.07% fue clasificado por el percentil 50, el 15.38% fue clasificado por

el percentil 75 y el 4.61% por el percentil 90. En la submuestra no hubo niños clasificados en el percentil 10. En cambio, el punto de corte cubano clasificó de la siguiente manera: 80% con bajo rendimiento intelectual y el 20% sin alteraciones.

Discusión

En la presente investigación, el análisis estadístico con una muestra mayor posibilitó apreciar un ajuste adecuado en las normativas cubanas del TMCR para niños en edad preescolar (de entre 4 y 6 años). En un estudio anterior, [Ramírez-Benítez et al. \(2015\)](#) ofrecieron un punto de corte para clasificar al niño en edad preescolar en dos grupos: niños con baja habilidad versus niños con habilidades típicas de la edad. Sin embargo, el resultado obtenido en este estudio ofrece una clasificación en cinco grupos: (1) niños con muy baja habilidad en el razonamiento analógico visual (percentil 10), (2) niños con baja habilidad (percentil 25), (3) niños con un desarrollo típico para la edad (percentil 50), (4) niños con alta habilidad (percentil 75) y (5) niños con habilidad superior (percentil 90).

La validez de los percentiles propuestos resultó adecuada cuando se la comparó con el punto de corte obtenido por [Ramírez-Benítez et al. \(2015\)](#). Además, el coeficiente de probabilidad diagnóstica indicó un descenso pequeño de la probabilidad diagnóstica de los percentiles obtenidos con respecto al punto de corte cubano.

En el primer resultado, la coincidencia diagnóstica entre los percentiles propuestos y el punto de corte cubano es del 69%. El 31% de los casos son considerados falsos negativos, pues los valores del percentil 50, obtenido por este estudio, son menores que los valores del punto de corte de [Ramírez-Benítez et al. \(2015\)](#). En consecuencia, un grupo de casos clasificados con bajo rendimiento

intelectual por el punto de corte cubano son clasificados como negativos por los percentiles propuestos. La caída o la elevación de los valores de los puntajes es uno de los principales problemas que encuentran los investigadores cuando ofrecen percentiles y los comparan con otros percentiles. Por ende, cada investigador debe estimar los beneficios y dificultades que la propuesta ofrece. Tal es el caso de los estudios de [Rossi-Casé et al. \(2015, 2016\)](#) cuando comparan los puntajes de los percentiles 50 en diferentes años en una población de adolescentes argentinos (1964, 2000 y 2015). En especial, el percentil 50 en 2000 para adolescentes de entre 13 y 15 años fue de 48 puntos, mientras que el percentil 50 en el 2015 fue de 45 puntos. En este caso, la comparación entre las dos muestras puede generar varios falsos negativos, lo cual no fue informado en el estudio, pues la intención de los autores era demostrar el efecto Flynn. De esta forma, se recomienda realizar un análisis de beneficios y dificultades diagnósticas cuando se ofrece una nueva normativa. En este caso, el CPD (coeficiente de probabilidad diagnóstica) es una alternativa para estimar el efecto de los percentiles en una determinada muestra.

El presente estudio, basándose en los resultados del CPD, reveló que los percentiles propuestos tienen un pequeño descenso en la probabilidad diagnóstica con respecto al punto de corte cubano. Desde esta perspectiva, y considerando los resultados de la coincidencia al clasificar, se recomienda precaución al utilizar los percentiles en la práctica, pues se requieren otros estudios para verificar sus beneficios y dificultades. No obstante, cuando se realizó la clasificación de ambos puntos de corte en la muestra de niños con dificultades en el neurodesarrollo (submuestra de [Ramírez-Benítez et al., 2015](#)), se pudo observar el impacto práctico de los percentiles propuestos. En especial, el punto de corte cubano indicó muy pocos casos de niños con rendimiento inte-

lectual típico para la edad (20%), lo cual es destacable, pues los problemas del neurodesarrollo tienen una incidencia en el desarrollo de la inteligencia, pero no de manera tan significativa como para diagnosticar o suponer alteración intelectual en el 80% de la muestra, menos aún si se considera que los autores tomaron casos de niños con alteraciones del neurodesarrollo sin discapacidad intelectual severa.

En este caso, los percentiles propuestos son más flexibles al clasificar los casos en la submuestra de Ramírez-Benítez et al. (2015), pues si bien los niños con problemas del neurodesarrollo, y en dependencia de otros factores clínicos y educativos, presentan bajo rendimiento intelectual (tal y como se indica en la Tabla 4, percentil 25; 46.92% de la submuestra), no aparece ningún niño clasificado por el percentil 10, lo cual indica que el bajo rendimiento intelectual en esta población no es tan significativo como se esperaba. Incluso se pudieron observar en los análisis algunos casos de niños clasificados en los percentiles 75 y 90.

Desde esta perspectiva, los percentiles propuestos pueden diferenciar a los niños con bajo rendimiento intelectual; ya sea por inmadurez en su neurodesarrollo debida a diversos factores (percentil 25) o por factores más relacionados con lo neurogenético, con muy bajos valores en el rendimiento intelectual (percentil 10). En igual sentido, estos pueden determinar si los niños no presentan dificultades en su rendimiento intelectual o si presentan valores superiores (percentiles 75 y 90) no esperados, aún cuando tengan alteraciones en el neurodesarrollo.

En relación con las propiedades psicométricas de la prueba Raven, en la literatura se han informado varias investigaciones que aseveran la validez y confiabilidad de la prueba, ya sea a través de la Teoría Clásica del Test (TCT; Flores-Mendoza, Widaman, Bacelar, & Lelé, 2014; Ibrahim & Kazem, 2013; Mansilla et al., 2012;

Van der Elst et al., 2013) así como en la Teoría de la Respuesta al Ítem (TRI; Escurra-Mayaute & Delgado-Vásquez, 2010; Sisto, Marín-Rueda, & Bartholomeu, 2006).

En el caso del estudio de Escurra-Mayaute y Delgado-Vásquez (2010), los investigadores verificaron las propiedades psicométricas de la prueba Raven según las dos teorías (TCT y TRI) en una población de jóvenes universitarios. Con respecto a la TCT, la prueba mostró una consistencia interna adecuada (coeficiente Kuder-Richardson 20 con un valor de .79). En relación con la TRI, los primeros 20 ítems de la prueba fueron fáciles para los jóvenes, mientras que los demás indicaron niveles de dificultad entre adecuado y difícil. El índice de discriminación reveló que todos los ítems tienen un nivel de discriminación entre bueno y adecuado, aunque el ítem 13 no mostró buena discriminación. De esta forma, los autores aseveraron que el análisis de ítems es una forma apropiada de determinar las debilidades y fortalezas de la prueba con respecto a la dificultad del ítem y su capacidad para discriminar la habilidad a evaluar en una población dada.

En este sentido, en la presente investigación se realizó un análisis de ítems para evaluar las propiedades psicométricas de la prueba Raven, así como saber si la mayoría de los ítems son fáciles o difíciles para los niños en edad preescolar y si estos discriminan entre niños con alta y baja habilidad.

En este caso, el coeficiente alfa de Cronbach cuando el ítem es eliminado indicó valores adecuados, entre .84 y .89, en la muestra general. En igual sentido los valores de alfa por edades fueron significativos. No obstante, los ítems con posibilidad de aumentar alfa, ya sea de manera general o por edades, se destacan por ser más fáciles o más difíciles para el niño preescolar, tal es el caso de los ítems 1, 4 y 32, los dos primeros por ser más fáciles y el último por ser más complejo. A pe-

sar del hecho de que haya varios ítems clasificados como difíciles o fáciles para los preescolares, la consistencia interna de la prueba es adecuada, tal y como se ha informado en varias investigaciones, ya sea a través de métodos test-retest (Mansilla et al., 2012; Rossi-Casé et al., 2015) o mediante el método de alfa de Cronbach (Ossa-Cornejo, Jara-Quezada, & Troncoso-San Martín, 2014; Ramírez-Benítez, Lorenzo-González, & Díaz-Bringas, 2013).

De igual manera, este argumento puede ser confirmado por la correlación ítem-total, pues los ítems aportan valores a la varianza total de la prueba de manera significativa. Los valores de las correlaciones ítem-total son de entre .33 y .56, es decir entre buenos y excelentes, según la clasificación de Ebel y Frisbie (1986).

De esta manera, se espera que la prueba identifique a niños con baja y alta habilidad en el razonamiento analógico. En efecto, el 100% de los ítems mostraron un índice de discriminación entre bueno y aceptable. Por ende, la prueba es adecuada para discriminar entre niños con baja y alta habilidad. En este sentido, la prueba Raven puede ser aplicada en edad preescolar, en especial en niños de entre 4 años y 5 meses y 6 años y 10 meses, pues en esa edad el niño puede resolver los problemas de matrices y al mismo tiempo es posible clasificar el caso de cada niño en uno de cinco grupos diagnósticos según el resultado obtenido.

A manera de conclusión, en los niños en edad preescolar es necesario conocer las potencialidades y debilidades del desarrollo, pues esto posibilita trazar una estrategia de estimulación acorde con las necesidades de cada niño. El TM-CR es una alternativa valiosa para el profesional de la educación y la salud, ya que es una prueba fácil de aplicar y sobre todo posibilita conocer el desarrollo del pensamiento del niño. De esta forma, los percentiles que ofrece este trabajo se establecen con la intención de identificar de ma-

nera temprana las dificultades en el pensamiento para organizar intervenciones educativas y evitar un posible fracaso escolar relacionado con dificultades en la inteligencia fluida. No obstante, se recomienda realizar otras investigaciones para validar los percentiles con una muestra en la que se controlen las alteraciones del neurodesarrollo, ya sean síndromes genéticos o síndromes multifactoriales. En igual sentido, sería oportuno evaluar una muestra de niños en edad preescolar y valorar el impacto de la clasificación en la edad escolar, pues se espera que los niños clasificados dentro del percentil 25 no tengan el mismo desempeño escolar que aquellos clasificados en el percentil 10. Además, se podría valorar el rendimiento de los falsos negativos obtenidos por los percentiles propuestos con respecto al punto de corte de Ramírez-Benítez et al. (2015). Sería de importancia determinar hasta qué punto se trata en estos casos de falsos negativos.

Conclusiones

1. El análisis de ítems (alfa cuando el ítem es eliminado, correlación ítem-total e índices de discriminación y dificultad) indicó que el Test de Matrices Progresivas Coloreadas de Raven presenta propiedades psicométricas adecuadas para aplicarse en niños en edad preescolar, en especial en las edades comprendidas entre 4 y 6 años.

2. El análisis de coincidencia entre los percentiles propuestos y el punto de corte cubano y el análisis con el coeficiente de probabilidad diagnóstica indicaron que los percentiles propuestos deben utilizarse en la práctica con precaución. No obstante, la clasificación de los percentiles en una muestra de niños con alteraciones en el neurodesarrollo es más flexible y operativa que el punto de corte que ofrecen Ramírez-Benítez et al. (2015).

Referencias

- Anstey, E. (1963). *Test de Dominós. Manual*. Buenos Aires: Paidós.
- Bandeira, D. R., Boccato-Alves, I. C., Giacometti, A. E., & Lorenzatto, L. (2004). Matrices Progressivas Coloridas de Raven - Escala Especial: Normas para Porto Alegre, RS. *Psicologia em Estudo*, 9(3), 479-486. doi: 10.1590/s1413-73722004000300016
- Bandeira, D. R., Costa, A., & Arteche, A. (2012). The Flynn effect in Brazil: Examining generational changes in the Draw-a-Person and in the Raven's Coloured Progressive Matrices. *Revista Latinoamericana de Psicología*, 44(3), 9-18. Recuperado de http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0120-05342012000300001&lng=en&tlng=en
- Brouwers, S. A., Van de Vijver, F. J. R., & Van Hemert, D. A. (2008). Variation in Raven's Progressive Matrices scores across time and place. *Learning and Individual Differences*, 19(3), 330-338. doi: 10.1016/j.lindif.2008.10.006
- Cairo-Valcárcel, E., Cairo-Martínez, E., Bouza, C., & Solozabal, T. P. (2000). Algunas características y posibilidades del Test de Matrices Progresivas de Raven. *Revista Cubana de Psicología*, 17(2), 95-105.
- Cattell, R. B., & Cattell, A. K. S. (1994). *Tests de Factor «g», Escalas 2 y 3*. Madrid: TEA Ediciones, S. A.
- Crocker, L., & Algina, J. (1986). *An introduction to classical and modern test theory*. New York: Holt, Rinehart & Winston.
- Crone, E., Wendelken, C., Van Leijenhorst, L., Honomichl, R., Christoff, K., & Bunge, S. (2009). Neurocognitive development of relational reasoning. *Developmental Science*, 12(1), 55-66. doi: 10.1111/j.1467-7687.2008.00743.x
- Ebel, R. L., & Frisbie, D. A. (1986). *Essentials of Education Measurement*. Englewood Cliffs, NJ: Prentice Hall.
- Escurra-Mayaute, L. M., & Delgado-Vásquez, A. E. (2010). Análisis psicométrico del Test de Matrices Progresivas Avanzadas de Raven mediante el modelo de tres parámetros de la teoría de la respuesta al ítem. *Persona*, 13, 71-97. Recuperado de <https://revistas.ultima.edu.pe/index.php/Persona>
- Fernández-Liporace, M., Ongarato, P., Saavedra, E., & Casullo, M. M. (2004). El Test de Matrices Progresivas, Escala General: Un análisis psicométrico. *Evaluar*, 4, 50-69. Recuperado de <https://revistas.unc.edu.ar/index.php/revaluar>
- Fernández-Nistal, M. T., & Mercado-Ibarra, S. M. (2014). Normative data of Raven's Coloured Progressive Matrices in Yaqui indigenous children. *Anuario de Psicología/The UB Journal of Psychology*, 44(3), 373-385. Recuperado de <http://revistes.ub.edu/index.php/Anuario-psicologia>
- Flores-Mendoza, C., Widaman, K. F., Bacelar, T. D., & Lelé, Á. J. (2014). Propriedades psicométricas do Raven Geral no contexto de Minas Gerais. *Arquivos Brasileiros de Psicologia*, 66(2), 1-16. Recuperado de http://pepsic.bvsalud.org/scielo.php?script=sci_issuetoc&pid=1809-526720140002&lng=pt&nr-m=iso
- Flynn, J. R. (2013). The "Flynn Effect" and Flynn's paradox. *Intelligence*, 41(6), 851-857. doi: 10.1016/j.intell.2013.06.014
- Flynn, J. R., & Rossi-Casé, L. (2012). IQ gains in Argentina between 1964 and 1998. *Intelligence*, 40(2), 145-150. doi: 10.1016/j.intell.2012.01.006
- IBM (2012). SPSS Statistics 21.0.0. [software de cómputo]. Disponible en <https://www-01.ibm.com/software/mx/analytics/spss/products/statistics>
- Ibrahim, A. M., & Kazem, A. M. (2013). Psychometric properties of scores from an embedded and independently-administered short form of the Raven's Advanced Progressive Matrices. *International Journal of Learning Management Systems*, 1(2), 25-35. doi: 10.12785/ijlms/010203
- Ivanovich, R., Forno, H., Durán, M. C., Hazbún-Game, J., Castro, J., & Ivanovich, C. (2000). Estudio de la capacidad intelectual (Test de Matrices Progresivas de Raven) en escolares chilenos de 5 a 18 años.

- Antecedentes generales, normas y recomendaciones. *Revista de Psicología General y Aplicada*, 53(1), 5-30. Recuperado de <https://dialnet.unirioja.es/servlet/articulo?codigo=2356766>
- Lázaro, J., Ostrosky, F., & Lozano, A. (2012). *Batería Neuropsicológica de Funciones Ejecutivas y Lóbulos Frontales-2*. Manual Moderno: México.
- Mansilla, C., Vásquez, D., & Estrada, C. (2012). Pertinencia normativa del Raven para la evaluación de población infantojuvenil socialmente vulnerable. *Terapia Psicológica*, 30(1), 73-80. doi: 10.4067/s0718-48082012000100007
- Ministerio de Salud Pública, Cuba. (2013). Dirección nacional de registros médicos y estadísticas de salud. Anuario estadístico de salud 2013. Recuperado de <files.sld.cu/dne/files/2014/05/anuario-2013-esp-e.pdf>
- Moratalla-Rodríguez, G. (2015). Lectura crítica de artículos de pruebas diagnósticas II: Análisis de resultados. *Radiología*, 57, 22-28. doi: 10.1016/j.rx.2014.11.004
- Neisser, U., Boodoo, G., Bouchard, T. J. Jr., Boykin, A. W., Brody, N., Ceci, S. J. ... Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, 51(2), 77-101. doi: 10.1037/0003-066x.51.2.77
- Ossa-Cornejo, C. J., Jara-Quezada, J., & Troncoso-San Martín, T. (2014). Validación Test de Matrices Progresivas de Raven Escala Coloreada, en escolares de la ciudad de Chillán. Memorias de Investigación. Recuperado de <http://repobib.ubiobio.cl/jspui/handle/123456789/204>
- Ramírez-Benítez, Y., Jiménez-Morales, R. M., & Díaz-Bringas, M. (2015). Matrices progresivas de Raven: Punto de corte para preescolares 4 - 6 años. *Evaluar*, 15, 123-133. Recuperado de <https://revistas.unc.edu.ar/index.php/revaluar>
- Ramírez-Benítez, Y., Lorenzo-González, G., & Díaz-Bringas, M. (2013). Matrices progresivas de Raven y su versión abreviada en la población escolar. *Revista Mexicana de Neurociencias*, 14(2), 63-67. Recuperado de <http://revmexneuroci.com/>
- Ramírez-Benítez, Y., Torres-Díaz, R., & Amor-Díaz, V. (2016). Contribución única de la inteligencia fluida y cristalizada en el rendimiento académico. *Revista Chilena de Neuropsicología*, 11(2), 1-5. Recuperado de <http://www.neurociencia.cl>
- Raven, J. C. (2004). *Test de Matrices Progresivas. Escala Coloreada. Cuaderno de Matrices / Series A, AB y B*. México: Editorial Paidós.
- Rossi-Casé, L., Neer, R., Lopetegui, S., Doná, S. M., Biganzoli, B., & Garzaniti, R. (2014). Matrices Progresivas de Raven: Efecto Flynn y actualización de baremos. *Revista de Psicología*, 23(2), 3-13. doi: 10.5354/0719-0581.2014.36144
- Rossi-Casé, L., Neer, R., Lopetegui, S., Doná, S. M., Biganzoli, B., & Garzaniti, R. (2015). Test de Raven, baremos argentinos para el rango 13-18 años y efecto Flynn. *Orientación y Sociedad*, 15, 1-24. Recuperado de http://www.scielo.org.ar/scielo.php?script=sci_arttext&pid=S1851-88932015000100005&lng=es&nrm=iso
- Rossi-Casé, L., Neer, R., Lopetegui, S., Doná, S. M., Biganzoli, B., & Garzaniti, R. (2016). Test de Raven: Actualización de baremos en adolescentes argentinos y análisis del efecto Flynn. *Revista Iberoamericana de Diagnóstico y Evaluación*, 42(2), 3-13. doi: 10.21865/RIDEP42_3
- Sisto, F. F., Marín-Rueda, F. J., & Bartholomeu, D. (2006). Estudo sobre a unidimensionalidade do Teste Matrices Progressivas Coloridas de Raven. *Psicologia: Reflexão e Crítica*, 19(1), 66-73. doi: 10.1590/S0102-79722006000100010
- Solovieva, Y., Loredó, D., Quintanar, L., & Lázaro, E. (2013). Caracterización neuropsicológica de una población infantil urbana a través de la Evaluación Neuropsicológica Infantil Puebla-Sevilla. *Pensamiento Psicológico*, 11(1), 83-98. Recuperado de <http://revistas.javerianacali.edu.co/index.php/pensamientopsicologico>
- Spearman, C. (1904). "General intelligence" objectively determined and measured. *The American Journal of Psychology*, 15(2), 201. doi: 10.2307/1412107
- Sundet, J. M., Barlaug, D. G., & Torjussen, T. M. (2004).

The end of the Flynn effect? A study of secular trends in mean intelligence test scores of Norwegian conscripts during half a century. *Intelligence*, 32, 349-362. doi: 10.1016/j.intell.2004.06.004

Teasdale, T. W., & Owen, D. R. (2007). Secular declines in cognitive test scores: A reversal of the Flynn effect. *Intelligence*, 36, 121-126. doi: 10.1016/j.intell.2007.01.007

Tenorio, M., Arango, P., Aparicio, A., Benavente, C., Thibaut, C., & Rosas, R. (2012). *Test de Evaluación Neuropsicológica Infantil*. Pontificia Universidad Católica de Chile.

Van der Elst, W., Ouweland, C., Van Rijn, P., Lee, N., Van Boxtel, M., & Jolles, J. (2013). The shortened Raven Standard Progressive Matrices: Item response theory-based psychometric analyses and normative data. *Assessment*, 20(1), 48-59. doi: 10.1177/1073191111415999

Wechsler, D. (2011). *Manual de Administración y Puntuación (WISC-IV)*. Buenos Aires: Paidós.

Wechsler, D. (2012). *Wechsler Preschool and Primary Scale of Intelligence (WPPSI-IV)* (4ª ed.). San Antonio, TX: Pearson.

Propiedades Psicométricas del Cuestionario de Adaptación para Adolescentes Bell en Población Bogotana

Psychometric Properties of the Bell Adolescent Adaptation Questionnaire in Bogota Population

Rocío del Pilar Gómez-Ramírez *¹, Guadalupe de la Iglesia²

1 - Universidad Nacional de Colombia, Colombia.

2 - Universidad de Buenos Aires, Argentina.

Introducción
Método
Análisis de datos
Resultados
Referencias

Recibido: 18/07/2017 **Revisado:** 29/07/2017 **Aceptado:** 12/08/2017

Resumen

El presente trabajo tuvo como objetivo establecer las propiedades psicométricas del Cuestionario de Adaptación para Adolescentes de Bell (*Bell Adjustment Inventory; 1934*), cuyo fin es operacionalizar el grado de adaptación en áreas relevantes del desarrollo adolescente. La muestra estuvo compuesta por 569 adolescentes entre 12 y 18 años residentes en la ciudad de Bogotá, Colombia. Se obtuvieron evidencias de validez de constructo por medio de un análisis factorial, cuyos resultados sugirieron mantener la estructura de cuatro dimensiones de adaptación: familiar, de salud, social y emocional. Se evaluó la consistencia interna de cada dimensión calculando los coeficientes alfa de Cronbach. Además, se estudiaron diferencias según sexo y edad. Los resultados obtenidos en cuanto a validez de constructo y consistencia interna son favorables e indican un buen funcionamiento psicométrico del instrumento.

Palabras clave: adaptación, adolescentes, propiedades psicométricas, pruebas psicológicas

Abstract

The aim of this study was to establish the psychometric properties of the Bell Adjustment Inventory for adolescents (*Bell, 1934*), an instrument that aims to operationalize the degree of adaptation in adolescents in different relevant areas of their development. The study was conducted on a sample of 569 adolescents between 12 and 18 years old residing in the city of Bogota, Colombia. The results showed evidence of construct validity through a factor analysis by keeping its structure of four adjustment dimensions that correspond with family, health, social and emotional. The internal consistency of each dimension was evaluated by calculating the Cronbach alphas. Also, differences depending on sex and age were studied. The results indicated a proper psychometric functioning of the instrument and evidences of construct validity were obtained as well as evidences of good internal consistency.

Keywords: adjustment, adolescent, psychometric properties, psychological test

*Correspondencia a: Rocío del Pilar Gómez Ramírez, Calle 39 BIS #29-90 Bogotá, Colombia. Te.: +573003066063. rdgomezr@unal.edu.co

Cómo citar este artículo: Gómez-Ramírez, R. del P., de la Iglesia, G. (2017). Propiedades psicométricas del Cuestionario de Adaptación para Adolescentes Bell en población bogotana. *Revista Evaluar*, 17(2), 47-66. Recuperado de <https://revistas.unc.edu.ar/index.php/revaluar>

Nota: El presente artículo es derivado de la investigación "Propiedades psicométricas del Cuestionario de Adaptación para Adolescentes de Bell en población bogotana" presentada como tesis de grado de maestría.

Introducción

La *adaptación* es reconocida como función comportamental universal y atributo interno de la conducta y está relacionada con el bienestar puesto que influye en la estabilidad emocional y personal (Jiménez & López-Zafra, 2011; Labajos-Alonso, 1996; Rodríguez, Jiménez, Fernández, & Godoy, 1999). Diferentes autores la definen como el proceso en el que un organismo vivo mantiene el equilibrio entre sus necesidades y las circunstancias que influyen en cómo estos logran satisfacerlas (Kulshrestha, 1979; Petersen & Crockett, 1985). En el proceso adaptativo el organismo se ajusta a un ambiente que se encarga de moldearlo, otorgándole rasgos que benefician su exploración.

A la vez, se ha descrito la adaptación como la respuesta a cambios o conflictos generados por factores internos (necesidades básicas como alimentación, descanso, estado de salud, entre otros) o externos (procesos de duelo, afrontamiento hacia distintos estímulos, ajuste social, entre otros) a partir del conocimiento de las necesidades propias (Bharvad, 2015; Cardenal-Hernández & Fierro-Bardaji, 2001; Del Bosque-Fuentes & Aragón-Borja, 2008; Holahan, Moos, & Schaefer, 1996; Macías, 2000).

De acuerdo con lo anterior, la adaptación está constituida por un doble proceso: la tendencia al ajuste de las personas a los propios deseos, gustos, preferencias y necesidades; y el ajuste de las propias conductas a las circunstancias de un entorno con normas y pautas en las relaciones interpersonales en el que se hallan insertos. En el doble proceso, el individuo no está aislado ni contrapuesto al mundo, sino por el contrario, siempre relacionado con él (Alcalay, Milicic, & Torretti, 2005; Fierro & Cardenal, 1996). Los mecanismos de ajuste que median la adaptación

parten de conductas y cogniciones propias del sujeto, proponiendo un acuerdo entre las características personales y las demandas del medio al que pertenece (García-Pérez & Magaz-Lagos, 2001; Gómez-Bustamante & Cogollo, 2010; Martínez-Ferrer, Amador-Muñoz, Moreno-Ruiz, & Musitu-Ochoa, 2011; Orte-Socias & March-Cerdá, 1996).

De esta forma se define la adaptación como un constructo relacional por ser un proceso activo e interactivo de ajuste a las necesidades y demandas propias y del entorno para hacer a este último habitable. La adaptación constituye, por ende, un hábito comportamental y un estado emocional (García-Pérez & Magaz-Lagos, 2001; Fierro, 1990, 1997; Fierro & Cardenal, 1996; Kulshrestha, 1979; Ochoa, 2012; Siverio-Eusebio & García-Hernández, 2007). Es así como la adecuada adaptación ocurre tanto por el reconocimiento de las emociones y habilidades personales como por el mantenimiento de relaciones personales, familiares y extrafamiliares adecuadas, lo cual permite explorar el mundo social y personal, así como conocer y apropiarse formas de comportamiento relacionadas con la idiosincrasia de un contexto.

Generalmente una adecuada adaptación se relaciona con altos niveles de autoestima, autoimagen y autoconcepto; adecuado desempeño académico y profesional y satisfacción vital (Kostanski & Gullone, 1998; Musitu & García 2001). De esta manera, las personas que logran una adecuada adaptación se caracterizarían por poseer altos niveles de autonomía e independencia con tendencias a la dinámica y vinculación relacional, satisfacción con su cotidianidad y adecuada resolución de dificultades (Fierro, 1997; Fierro & Cardenal, 1996; Kulshrestha, 1979; Ochoa, 2012; Siverio-Eusebio & García-Hernández, 2007).

La adaptación en la adolescencia

La adaptación se da en todas las etapas de desarrollo y determina la calidad de vida a nivel biológico, emocional, psicológico y social. Frente a esto se ha descrito que el logro de la adaptación en la adolescencia es fundamental (Carvajal-Corzo, 1993; Gómez-Bustamante & Cogollo, 2010; Labajos-Alonso, 1996). Se entiende por adolescencia una etapa del ciclo vital del desarrollo entre la infancia y la adultez donde se destacan la apropiación de una identidad y la vinculación afectiva, que resultan en la inserción en un grupo social de pertenencia (Carvajal-Corzo, 1993). Es evidente que la adaptación es de especial importancia en esta etapa en la que se da una serie de inmensos cambios inicialmente biológicos y posteriormente psicológicos.

El estudio de la adaptación adolescente se enfoca en la estabilidad personal, la autopercepción emocional y del estado físico y las interacciones sociales (Coleman & Hendry, 1999; González, De los Ríos, & Viveros, 2015; Lynch, Norem, & Gergen, 1981). En esta etapa, la adaptación se consolida a través de variables mediadoras, a saber: *biológicas*, que representan el inicio de la pubertad afín a los cambios corporales, el temperamento y estados de salud en general y las cuales se desprenden del desarrollo orgánico y físico-psicosexual; *psicológicas*, relacionadas con el estado emocional, inteligencia, autopercepción, desarrollo del ego, desarrollo moral y desarrollo cognoscitivo; y *sociales*, en las que se denotan las nuevas configuraciones en la relaciones de pares y la entrada al mundo social y cultural (Alcalay et al., 2005; Coleman & Hendry, 1999; Fierro, 1990).

La evidencia empírica a la fecha sugiere que la edad es un factor relevante asociado a la adaptación. Esto es debido a que la etapa adolescente se divide en tres subetapas según rangos de edad.

Cada una de estas cumple un papel importante en el desarrollo, ya que unifica características fisiológicas, reacciones y conductas comunes presentadas en los adolescentes de acuerdo con el proceso de desarrollo (Carvajal-Corzo, 1993; Ros-Rahola et al., 2006). En la Tabla 1 se encuentran condensadas las características de cada una de las etapas de la adolescencia según Carvajal-Corzo (1993).

Asimismo, el sexo es otro factor relevante en la adaptación adolescente. En estudios referidos a las diferencias según sexo, por ejemplo, se ha descrito que las mujeres adolescentes muestran un mayor grado de inadaptación que los varones en un ámbito educativo, ya que estas se perciben como víctimas frente a la hostilidad de pares, mientras que los hombres tienden a la formación de grupos, observándose a sí mismos como ajustados al entorno (Funes, Lupiáñez, & Humphreys, 2010). Estos hallazgos coinciden con los descritos por Aragón-Borja y Bosques (2012), según los cuales los hombres adolescentes refieren mayores índices de adaptación debido a la poca importancia que le dan a los cambios físicos y, por lo mismo, a la autoimagen percibida en la inmersión en grupos, en los que la comunicación con pares trabaja como factor relevante para los procesos de adaptación.

Sin embargo, otros estudios señalan mayores índices de adaptación en las mujeres adolescentes que en los hombres en lo que refiere a la necesidad de ajustarse a los cambios que implica esta etapa, especialmente aquellos vinculados con los cambios corporales. Para estos autores el hecho de que el cuerpo femenino se desarrolle a mayor velocidad y este desarrollo sea más evidente que en los hombres propende al afán por ajustarse a las presiones, a los patrones sociales y al grupo de pares. Asimismo, las mujeres parecen recibir la etapa adolescente con reflexiones sobre lo que deviene de la misma, al contrario que los hombres (Bharvad, 2015; Del Bosque-Fuentes & Aragón-

Tabla 1

Etapas del período adolescente (Carvajal-Corzo, 1993).

Adolescencia Primaria (12 - 13 años)	<p>Caracterizada por los primeros cambios físicos, a saber: a) el aumento de la estatura, b) el aumento del peso corporal y de fuerza muscular, c) el cambio de voz y d) la aparición de los caracteres sexuales secundarios.</p> <p>Se destaca la preocupación de los adolescentes por los cambios corporales y de autoconcepto e imagen, por lo que surgen frecuentemente las comparaciones frente a pares del mismo sexo. A la par sucede el incremento de integración y contacto con personas del mismo sexo, se construyen los primeros índices de búsqueda de identidad a través de la pertenencia de grupos de amigos y lazos fraternales, así como las primeras sensaciones eróticas y la necesidad de intimidad.</p> <p>Se dan las primeras manifestaciones de ambivalencia emocional, especialmente la necesidad de independencia y nostalgia por la dependencia llevada hasta el momento con los padres.</p> <p>En cuanto al nivel cognitivo, en esta etapa se produce la evolución del pensamiento abstracto, la idealización de las vocaciones; se da menos control de los impulsos y mayor entrada al mundo de las fantasías.</p>
Adolescencia Nuclear (14 a 15 años)	<p>Se caracteriza por la remodelación morfológica y la conformación de la identidad sexual. Aumenta la atracción por el sexo opuesto, por lo que surgen actividades en pro del cuidado del cuerpo y la figura.</p> <p>En el área cognitiva se inicia el pensamiento abstracto y reaparecen manifestaciones egocéntricas.</p> <p>Se incrementan los conflictos con padres y autoridades, ya que se anhela la trasgresión de las normas, se da mayor fuerza a las opiniones propias y aumenta el sentido de independencia y la rebeldía.</p> <p>Se busca la independencia, el adolescente se percibe a sí mismo como poderoso y fuerte, características que llevan a comportamientos agresivos y en ocasiones peligrosos, como el consumo de drogas, el fumar, la ingesta de alcohol, entre otros.</p>
Adolescencia tardía (16 a 18 años)	<p>Esta marca el final del período adolescente. En ambos sexos se alcanzan las características del adulto en cuanto a funciones plenas para la sexualidad y reproducción. Se acepta la propia imagen y se establecen preferencias de amigos, música, forma de vestir, entre otras.</p> <p>En el área social se vuelve a apreciar el valor y consejo de los padres y se da la creciente integración en el medio social. Se establecen las relaciones íntimas, como la formación de parejas y amistades cercanas, relaciones que se caracterizan especialmente por el compartimiento de experiencias.</p> <p>Se presenta la maduración cognitiva y la convivencia emocional en donde se logra el control de impulsos emocionales, la capacidad para comprometerse y establecer límites y la concreción de valores y objetos personales.</p>

Borja, 2008; Alonso-Fernández, 2005).

En cuanto la edad como factor adaptativo, Cardenal-Hernández y Fierro-Bardaji (2001) no encontraron diferencias significativas entre la adolescencia temprana, la nuclear y la tardía en cuanto a la adaptación evaluada entre hombres y mujeres. Sin embargo, describen que los adolescentes entre 16 y 18 años presentan mayores ín-

lices de ajuste orientados a la protección de los demás, la reflexión sobre proyectos de vida y el conformismo. Se afirma que esto ocurre así debido a que en la adolescencia tardía los jóvenes se acercan al final del período, entrando al proceso de ajuste adulto. Asimismo, Aragón-Borja y Bosques (2012) hallaron que a mayor edad, mayores índices de adaptación adolescente, es decir, los

jóvenes establecidos en la fase de adolescencia tardía tienden a estar más adaptados. Por su parte, los jóvenes en el período de adolescencia nuclear presentan mayores índices de inadaptación por estar atravesando mayores cambios durante esta fase y tienden a la introversión y al poco reconocimiento de sí mismos.

Operacionalizaciones de la adaptación

A pesar de la importancia de la adaptación en la adolescencia, son escasos los instrumentos construidos para su medición en esta etapa. En la actualidad las escalas desarrolladas no se enfocan exclusivamente en la medición de la adaptación sino que valoran subaspectos relacionados o constructos similares entre los que están: la evaluación de la autoestima, para la cual se han desarrollado el Cuestionario de Autoestima Global de Rosenberg (1966) y el Inventario de Depresión Infantil (CDI) de Kovacs (2004), la evaluación de autoeficacia, en la que se destaca la Escala de Autoeficacia General de Baessler y Schwarzer (1996), la evaluación de la satisfacción con la vida, para la que se encuentran la Escala de Satisfacción con la Vida de Diener, Emmons, Larsen y Griffin (1985), su estandarización española realizada por Atienza-González, Moreno-Sigüenza y Balaguer-Solá (2000) y la Escala Multidimensional de Satisfacción con la Vida de Huebner (1991); la evaluación del clima escolar con la Escala de Clima Escolar (Moos, 1974), la adaptación española de Fernández-Ballesteros y Sierra (1984); y la evaluación de las habilidades sociales, para la cual existen la Escala de Actitudes y Estrategias Cognitivas Sociales (AECS) de Moraleda, González-Galán y García-Gallo (1998) y la Escala Multidimensional de Expresión Social (EMES) de Caballo y Ortega (1989), adaptada por Gutiérrez (2000). Estos instrumentos, a pesar de ser relevantes por estu-

diar aspectos de la conducta y eficacia adolescente, no brindan una medida general de adaptación. Por su parte, el Cuestionario de Adaptación para Adolescentes de Bell (CAAB; Bell, 1934) es un instrumento específicamente diseñado para evaluar la adaptación adolescente.

Cuestionario de Adaptación para Adolescentes de Bell

El Cuestionario de Adaptación para Adolescentes de Bell (CAAB; Bell, 1934) evalúa el estado general de adaptación adolescente de acuerdo con el modelo de adaptación propuesto por su autor según el cual son cuatro las dimensiones representativas del proceso de adaptación adolescente: familiar, social, de la salud y emocional. Estas dimensiones no son excluyentes entre sí, por el contrario interactúan, propiciando el referente de estabilidad emocional y el adecuado desenvolvimiento en un contexto determinado.

La adaptación *familiar* indica el grado de satisfacción del adolescente con los padres, principalmente, y su familia en general (Bell, 1934). La adaptación *social* se refiere al establecimiento y mantenimiento de relaciones personales afectivas adecuadas y el dominio de tareas relacionadas con la edad. Un buen desempeño social proporciona el sentido de pertenencia al contexto cultural, lo cual hace que el adolescente perciba satisfacción general dentro de las esferas sociales que lo rodean (Bell, 1934; Cerdá, 1987; Jiménez & López-Zafra 2011; Valverde, 1988). La adaptación *de la salud* es definida como un estado completo de adecuación física, lo que implica no sólo la ausencia de afección o enfermedad, sino la total cobertura de las necesidades fundamentales de manera sanitaria y nutricional. Este tipo de adaptación refiere a la adecuada condición de salud física, orgánica y corporal (Palmero & Fernández, 1998). Final-

mente, la adaptación *emocional* se relaciona con el grado de reconocimiento por parte del adolescente tanto de sus propios sentimientos como de los de los demás para manejarlos y controlarlos adecuadamente (Bell, 1934; Lynch et al., 1981; Oliva-Delgado & Parra-Jiménez, 2004; Valverde, 1988).

Los primeros trabajos psicométricos del cuestionario fueron realizados por Bell (1934), quien seleccionó cada una de las dimensiones del cuestionario partiendo del grado en que diferenciaban entre el 15% de individuos superior y el 15% inferior en una distribución de puntuación obtenida en la aplicación piloto del cuestionario de adaptación. Únicamente los ítems que diferenciaban claramente los grupos extremos se incluyeron en la construcción del cuestionario para la versión en inglés.

Para la validación, el cuestionario fue administrado a un grupo de estudiantes que además fueron calificados en los correspondientes sectores de adaptación por directores y psicólogos escolares. La escala de adaptación *social* se validó mediante la correlación Pearson con la escala adaptación social del Cuestionario de Ascendencia-Sumisión de Allport y Allport (1939) y B4-D del cuestionario de personalidad de Bernreuter (B4-D; Bernreuter, 1935), y la escala adaptación *emocional* y la puntuación total se validaron buscando también la correlación con el Personality Schedule de Thurstone (Thurstone, 1930). Los coeficientes de correlación obtenidos fueron: Cuestionario de Ascendencia-Sumisión y adaptación social (hombres) = .72; Cuestionario de Ascendencia-Sumisión y adaptación social (mujeres) = .81; Personality Schedule de Thurstone y adaptación emocional = .93, Personality Schedule de Thurstone y puntuación total = .94; Cuestionario de personalidad de Bernreuter y adaptación social = .90.

Para corroborar la confiabilidad se calcula-

ron coeficientes ρ para cada una de las dimensiones del cuestionario y del puntaje global a través de la fórmula de Spearman-Brown. La puntuación fue significativa: adaptación familiar, $\rho = .89$, adaptación de la salud, $\rho = .80$; adaptación social, $\rho = .89$; adaptación emocional, $\rho = .85$ y puntuación total, $\rho = .93$

Finalmente Bell (1934) para la validación del instrumento, realizó intercorrelaciones para las cuatro escalas del cuestionario obteniendo: Familia y Salud = .43; Familia y Social = .04; Familia y Emocional = .38; Salud y Social = .24; Salud y Emocional = .53 y, Social y Emocional = .47

En la versión española adaptada por Cerdá (1987) se modificó el texto de 14 ítems en concordancia con el contexto español, para mantener igual número de ítems del cuestionario original. Los coeficientes de confiabilidad son de $\alpha = .89$ para la escala de adaptación familiar, $\alpha = .80$ para la escala de adaptación en salud; $\alpha = .89$ para la escala de adaptación social y $\alpha = .84$ para la escala de adaptación emocional. En cuanto a la validez de contenido y de constructo, se analizó cada uno de los ítems en cuanto a conceptos de acuerdo con el contenido de cada una de las dimensiones. Se utilizó la matriz de correlación inter-subescala y se detectó una correlación negativa alta entre la adaptación y no adaptación, es decir, adecuados índices psicométricos para la versión española.

En un estudio con una versión mexicana del instrumento se trabajó con una muestra de 562 estudiantes de la Ciudad de México. Para esta adaptación se utilizaron las mismas evaluaciones que para la adaptación española, calculando el índice de confiabilidad para cada una de las escalas ($\alpha = .80$, para el área familiar, $\alpha = .75$ para el área de salud, $\alpha = .71$ para el área social, $\alpha = .85$ para el área emocional y $\alpha = .92$ para el total del cuestionario). En este trabajo, la escala resultante mantuvo la misma estructura de la adaptación española (Del Bosque-Fuentes & Aragón-Borja, 2008).

En países como Perú, Ecuador y Chile se ha utilizado la versión española del CAAB para diferentes estudios, pero no se llegó a adaptar el cuestionario (Alulema-Dávila & Tintin-Rea, 2014; Minda-Mina, 2011; Villadurna-Ríos, 2013).

En Colombia aún no se ha realizado la adaptación de este instrumento. Existen investigaciones que analizaron diversas problemáticas en población adolescente, que generalmente se centraron en sus problemas de conducta (Alba, 2007; Alonso-García & Román-Sánchez, 2005; Aylla-Velázquez, Pedroza-Cabrera, Morales-Chainé, Chaparro-Caso-López, & Barragán-Torres, 2002; Departamento Nacional de Estadística [DANE], 2005; Flores, 1992; Gómez-Restrepo, 2005; Profamilia, 2016; Programa Rumbos, 2001; Rabazo-Méndez, 1999; Rodríguez-Mayoral, Martínez-Arias, Díaz-Aguado, & Moretin, 2008; Suarez & Krauskopf, 1995; Vinaccia, Quiceno, & Moreno-San-Pedro, 2007). Aunque estas investigaciones son sumamente importantes, dejan de lado un aspecto fundamental de la adolescencia: la capacidad de adaptación del adolescente. Para poder realizar investigaciones que incluyan esta variable, es primordial contar con una medida adecuada del constructo. Es por ello que en este trabajo se propone la adaptación psicométrica de la versión española del Cuestionario de Adaptación para Adolescentes de Bell (CAAB; Bell, 1934) realizada por Cerdá (1987), para su uso con población adolescente de la ciudad de Bogotá, Colombia. La adaptación del CAAB posibilitará contar con una herramienta adecuada para caracterizar la adaptación de los adolescentes colombianos de la ciudad de Bogotá y su estudio en relación con otras variables relevantes. Esta información podrá ser utilizada por los actores implicados en los ámbitos educativos y sociales en los que los adolescentes se ven insertos –docentes, psicólogos, psicopedagogos, familiares, grupos de pares, entre otros– para idear interven-

ciones, tomar decisiones y actuar utilizando información generada desde una medida de adaptación apropiada.

Método

Participantes

Se trabajó con una muestra no probabilística constituida por 569 adolescentes, con edades entre 12 y 18 años ($M = 15.35$; $DE = 2.01$). Los participantes del estudio fueron estudiantes de escuelas primarias y secundarias y universidades públicas y privadas en la ciudad de Bogotá. El 53.8% de la muestra fueron mujeres ($n = 306$) y el 46.2% hombres ($n = 262$). En cuanto al estado civil de los participantes, el 83% ($n = 472$) informó ser soltero, el 16.3% ($n = 93$) tenía una relación y sólo el 0.7% ($n = 4$) vivía con su pareja en unión libre.

Con el fin de cumplir con los objetivos del estudio, se conformaron grupos de edad de acuerdo a la subetapa de desarrollo en la adolescencia (ver Tabla 1). Ciento veinte participantes se encontraban en la adolescencia temprana (21.1%), 168 participantes se encontraban en la adolescencia nuclear (29.5%) y 281 se encontraban en la adolescencia tardía (49.4%).

Tabla 2

Distribución de la muestra agrupada por edad según subetapa de adolescencia.

	Frecuencia	Porcentaje
Adolescencia Temprana (12 -13 años)	120	21.1
Adolescencia Nuclear (13-14 años)	168	29.5
Adolescencia Tardía (16-18 años)	281	49.4
Total	569	100

En cuanto al nivel socioeconómico, en Colombia se utiliza una clasificación de acuerdo a niveles de estratificación. Los estratos socioeconómicos en los que se pueden clasificar las viviendas y/o predios son seis, a saber: 1 = bajo-bajo, 2 = bajo, 3 = medio-bajo, 4 = medio, 5 = medio-alto y 6 = alto. Los estratos 1, 2 y 3 corresponden a los niveles bajos, a los que pertenecen las personas con menores recursos, quienes son beneficiarios de subsidios en los servicios públicos domiciliarios. Por su parte, los estratos 5 y 6 corresponden a estratos altos a los que pertenecen los ciudadanos con mayores recursos económicos, quienes a la vez deben pagar contribuciones sobre el valor de los servicios públicos domiciliarios. El estrato 4 no es beneficiario de subsidios ni debe pagar sobrecostos, paga exactamente el valor que las empresas definan como costo de prestación del servicio. La clasificación en estratos es una aproximación a la diferencia socioeconómica jerarquizada, es decir, de pobreza a riqueza y viceversa. Esta clasificación se basa en la Ley 142 de 1994 (régimen de servicios públicos domiciliarios) en la que se dictamina que las características físicas se utilizan como base para asignar los estratos a los inmuebles residenciales de la población colombiana. Se deben estratificar los inmuebles residenciales y no los hogares (artículo 101.2, Ley 142; DANE, 2005). De acuerdo con esta clasificación, la mayoría de la muestra estudiada se encontraba en el estrato 3 (45.3%; n = 258), seguido por el estrato 2 (27.4%; n = 156), el estrato 4 (20,4%; n = 116), el estrato 1 (4.2%; n = 24), el estrato 5 (2.5%; n =14) y en menor porcentaje el estrato 6 (0.2%; n = 1).

Materiales

Encuesta sociodemográfica. La misma fue diseñada ad hoc para esta investigación. Mediante ella

se indagó acerca de la edad, sexo, estado civil, nivel socioeconómico y grado académico de los adolescentes.

Cuestionario de Adaptación para Adolescentes de Bell. Se utilizó la adaptación española para elaborar una versión colombiana (Cerdá, 1987). Este test evalúa el nivel de adaptación de adolescentes de entre los 12 y 18 años de edad a través de 140 ítems agrupados en cuatro dimensiones de adaptación: familiar, de la salud, social y emocional. Para dar respuesta al cuestionario los participantes respondieron a los ítems, que constan de afirmaciones según la correspondencia o no de cada una con su forma de ser y de pensar o con sus percepciones e intereses. Las opciones de respuesta del cuestionario son *SI*, *NO* y *?*, este último utilizado en caso de no encontrarse seguro con su respuesta. Previo a la calificación del test deben invertirse algunos ítems cuya formulación va en sentido opuesto. La suma de las respuestas clave permite obtener el nivel de adaptación para cada una de las áreas de adaptación. A mayor puntaje en la puntuación bruta de cada escala, menor grado de adaptación.

Procedimiento

La versión preliminar del CAAB fue sometida a juicio experto, estudio piloto y administrada a una muestra descriptiva con el fin de analizar la redacción de cada uno de los ítems en el contexto colombiano y el poder discriminativo de los ítems e identificar la estructura factorial. La administración del cuestionario se realizó de manera colectiva. Se procuró la aplicación del instrumento en condiciones de adecuada iluminación, ventilación y bajos niveles de ruido para que los participantes respondieran motivados y, por consiguiente, de forma honesta. Se solicitó un consentimiento informado, asegurando la confidencialidad de los

datos y la posibilidad de una devolución individual de los resultados obtenidos en caso de ser solicitada. Asimismo, se informó la posibilidad de no realizar el cuestionario o dejar de responderlo en el momento en que los participantes lo consideraran pertinente.

Análisis de datos

Para obtener evidencias de validez de constructo se realizó un análisis de componentes principales. Se verificó la adecuación de la muestra mediante el estadístico Kaiser-Meyer-Olkin (KMO) y el test de esfericidad de Bartlett. Luego se analizó confiabilidad mediante el coeficiente alfa de Cronbach para cada una de las escalas. Posteriormente, se realizaron análisis de diferencias entre grupos para observar las diferencias entre las variables de sexo, y edad, para lo cual se analizaron los estadísticos descriptivos (*t* de Student) y se llevó a cabo un ANOVA.

Finalmente se realizó la construcción de normas agrupando los baremos según las etapas de la adolescencia, para lo cual se emplearon los puntajes percentiles y los puntajes *Z*.

Resultados

Juicio experto

El juicio experto fue realizado por cinco jueces quienes evaluaron la consigna, el formato de respuesta y el contenido de los ítems. Como resultado se modificó parte de la consigna para mayor comprensión de los participantes. En cuanto al formato, se determinó que para la aplicación del instrumento las opciones de respuesta se ubicarían al final de cada uno de los ítems con el fin de facilitar la selección de las opciones de respuesta una vez terminada la lectura del ítem, ya que de

otra forma podrían funcionar como un distractor en la lectura. Se realizaron modificaciones en la redacción de los ítems en relación con la versión española, adecuándose las oraciones a los modismos colombianos y al vocabulario adolescente.

Con el fin de calcular los coeficientes Aiken, se solicitó a los jueces indicar la pertenencia de cada ítem respecto a cada subdimensión adaptativa para así valorar el grado de acuerdo. Los resultados arrojaron en general valores adecuados, excepto para el ítem 42 ($V = .0$), el ítem 106 ($V = .0$) y el ítem 120 ($V = .4$), que presentaron valores Aiken menores que $.60$, siendo retirados de los análisis posteriores teniendo en cuenta la no pertenencia de los ítems con cada una de las escalas.

Estudio Piloto

Con la versión post-juicio experto del instrumento se llevó a cabo un estudio piloto con 17 adolescentes quienes completaron y evaluaron la estructura del test en cuanto a la consigna, el formato de respuesta y el contenido de los ítems. Diversas observaciones promovieron modificaciones en el instrumento respecto a la descripción de la consigna en donde se aclaró a qué hacían referencia las respuestas *SI*, *NO* y *?* y la redacción de algunos de los ítems los cuales se adaptaron a la edad de los participantes y el contexto colombiano. No surgieron modificaciones en el formato del test.

Análisis de ítems: Discriminación y homogeneidad

Para evaluar la capacidad discriminativa de los ítems para cada una de las dimensiones que componen el instrumento como primer paso se suprimieron aquellos ítems en los cuales la frecuencia fuera mayor que 70% en una de las tres

opciones de respuesta (Rodríguez-Sabiote, Gallardo-Vigil, Pozo-Lorente, & Gutiérrez-Pérez, 2006). Se examinó la normalidad univariante mediante los estadísticos de contraste (zG) para los valores de asimetría y curtosis con sus correspondientes errores típicos, esperando el rango de $\pm 1,96$ (Lévy-Mangin & Varela-Mallou, 2006). Luego, se calcularon los índices de homogeneidad corregidos (IHC) por subescala teniendo en cuenta aquellos índices con valores inferiores a .30 (Martínez-Arias, 1995).

Análisis factorial

Los 137 ítems fueron sometidos a un análisis de componentes principales con rotación seleccionada varimax. Para comenzar se estudió el gráfico de sedimentación y los autovalores. Con el fin de intentar replicar la estructura de la prueba creada por Bell se forzó la extracción de cuatro factores. Se eliminaron elementos con cargas inferiores a .40 o cargas dobles y se logró aislar una estructura de 24 ítems que explicaba el 36.77% de la varianza ($\chi^2 = 1823,71$; $gl = 276$; $p = .0001$; $KMO = .76$). La estructura puede visualizarse en la Tabla 3. El Factor 1 corresponde con los ítems de la escala *familiar*, el Factor 2 corresponde con los ítems de la escala *de salud*, el Factor 3 corresponde con los ítems de la escala *social* y por último, el Factor 4 corresponde con los ítems de la escala *emocional*. También se calcularon los coeficientes alfa de Cronbach para valoración de consistencia interna. Tal como puede verse también en la Tabla 3, se verificaron valores adecuados para todas las escalas a excepción de la escala *de salud*.

Diferencias según sexo y edad

Con el fin de conocer diferencias estadísticamente significativas en las escalas de adaptación familiar, de salud, social y emocional, a partir de la variable *sexo* se utilizó una prueba t de Student para muestras independientes.

Los resultados muestran una diferencia estadísticamente significativa en las dimensiones de adaptación familiar, $t_{(567)} = -3.13$, $p = .002$ y adaptación emocional $t_{(567)} = -8.10$, $p < .001$. De acuerdo con lo anterior se encuentra que las mujeres presentan menores índices de adaptación familiar que los hombres, $M_V = 1.99$ vs. $M_M = 2.44$ y también menores índices de adaptación emocional, $M_V = 1.98$ vs. $M_M = 3.19$. Debe recordarse que en esta escala a mayor puntaje, menor adaptación. Al mismo tiempo, no se hallaron diferencias estadísticamente significativas en las subescalas de adaptación de salud y social ($p = .05$).

A continuación, para verificar si existían diferencias estadísticamente significativas en las escalas de adaptación familiar, de salud, social y emocional según el grupo de edad, se calcularon pruebas ANOVA one-way. Se encontró una diferencia estadísticamente significativa en relación con la adaptación familiar, $F_{(2,566)} = 3.41$, $p < .001$ y en relación con la adaptación emocional, $F_{(2,566)} = 11.08$, $p = .001$. En el caso de la adaptación familiar, la prueba post-hoc de Bonferroni indicó que existían diferencias significativas ($p = .05$) entre la adolescencia nuclear ($M = 2.52$) y la tardía ($M = 2.13$). Para la adaptación emocional la prueba post-hoc indicó que existían diferencias estadísticamente significativas entre todos los pares de rangos de edad. Los que pertenecían a la adolescencia tardía mostraron más adaptación ($M = 2.57$) que la adolescencia nuclear ($M = 3.12$). Así mismo, los adolescentes que se encontraron en la adaptación temprana, mostraron mayor adaptación ($M = 2.09$) que los adolescentes en etapa nu-

Tabla 3

Estructura Factorial del CAAB.

Ítem	Factor			
	1	2	3	4
Ítem 35. ¿Lloras con facilidad?				.56
Ítem 57. ¿Te desanimas fácilmente?				.65
Ítem 113. ¿Te encuentras a menudo como si sintieras por dentro una especie de nerviosismo?				.57
Ítem 123. Tras haber sufrido una situación humillante, ¿te quedas muy preocupado(a) durante bastante tiempo?				.64
Ítem 128. ¿Tienes altibajos en tu humor sin una razón aparente?				.63
Ítem 132. ¿Te dan a menudo las ideas tantas vueltas en la cabeza que te impiden dormir normalmente?				.53
Ítem 33. ¿Has sufrido alguna herida o lesión importante a consecuencia de algún accidente?		.50		
Ítem 38. ¿Has tenido alguna operación quirúrgica importante?		.52		
Ítem 50. ¿Has tenido dos o más enfermedades importantes durante los últimos diez años? (enfermedades graves o que te hayan obligado a guardar cama durante varias semanas)		.62		
Ítem 69. ¿Has necesitado que te visiten los médicos con bastante frecuencia?		.65		
Ítem 124. ¿Faltas a clase con cierta frecuencia por estar enfermo(a)?		.65		
Ítem 21. ¿Te parece que en tu casa debía de existir un poco más de comprensión y de afecto?	.65			
Ítem 51. ¿Estás casi siempre en desacuerdo con la forma en que enfoca las cuestiones de casa alguno de tus padres?	.64			
Ítem 67. ¿Te parece que tus padres han sido o son demasiado severos contigo?	.63			
Ítem 72. ¿Alguno de tus padres tiene costumbres o modales que te desagradan?	.53			
Ítem 82. ¿Te ves obligado(a) de vez en cuando a callarte o a marcharte a otra habitación con la finalidad de que haya paz y tranquilidad en tu hogar?	.46			
Ítem 134. ¿Tienes frecuentemente la impresión de que tus padres no te comprenden bien?	.61			
Ítem 8. Cuando estás en una reunión o con un grupo de amigos, ¿te gusta ser la encargada de presentar unos a otros? (INVERSO)			.56	
Ítem 15. ¿Has tomado a veces la iniciativa para animar alguna reunión aburrida? (INVERSO)			.60	
Ítem 49. ¿Te resulta difícil entablar conversación con una persona que te acaban de presentar?			.49	
Ítem 70. ¿Te resulta muy difícil hablar en público?			.52	
Ítem 96. ¿Haces amigos(as) con facilidad? (INVERSO)			.57	
Ítem 127. ¿Te gusta asistir a fiestas y reuniones a las que acuden muchas personas? (INVERSO)			.54	
Ítem 130. ¿Te sientes seguro(a) de ti mismo(a) cuando tienes que hablar en clase? (INVERSO)			.46	
Alfa de Cronbach	.66	.49	.62	.70
%Varianza Explicada	9.36	7.21	9.25	10.95
%Varianza Total Explicada		36.77		

clear. No se hallaron diferencias estadísticamente significativas en la adaptación de salud y la adaptación social en relación con el grupo de edad del adolescente ($p = .05$).

Finalmente, se calcularon los percentiles para las puntuaciones directas de las dimensiones familiar, de salud, social y emocional, de acuerdo con cada una de las subetapas de desarrollo adolescente.

Tabla 4
Baremos del Cuestionario de Adaptación Adolescente en población bogotana.

		Subescalas			
		Fami- liar	Salud	Social	Emocio- nal
De 12 a	N	120	120	120	120
13 años	Media	2.07	.86	2.86	2.09
	Mediana	2.00	1.00	3.00	2.00
	Moda	1	0	3	1
	D.E	1.719	.946	1.726	1.670
	Mínimo	0	0	0	0
	Máximo	6	4	7	6
Percentiles					
	10	0		0	0
	20			1	
	30				
	40	1	0	2	1
	50				
	60	2		3	2
	70	3	1		3
	80	4		4	
	90	≥ 5	≥ 2	≥ 5	≥ 4

		Subescalas			
		Fami- liar	Salud	Social	Emocio- nal
De 14 a	N	168	168	168	168
15 años	Media	2.52	.85	2.91	3.12
	Mediana	2.00	.00	3.00	3.00
	Moda	2	0	2	5
	D.E	1.775	1.087	1.847	1.932
	Mínimo	0	0	0	0
	Máximo	6	5	7	6
Percentiles					
	10	0			0
	20			≤ 1	1
	30	1			2
	40			2	
	50	2	0		3
	60	3		3	4
	70		1	4	
	80	4		5	5
	90	≥ 5	≥ 2	≥ 6	≥ 6

De 16 a	N	281	281	281	281
18 años	Media	2.13	.71	2.66	2.57
	Mediana	2	0	2	2
	Moda	0	0	2	0
	D.E	1.67	.997	1.898	1.875
	Mínimo	0	0	0	0
	Máximo	6	5	7	6
Percentiles					
	10			0	0
	20	0			
	30	1		1	1
	40				
	50	2	0	2	2
	60			3	3
	70	3			
	80	4	1	4	4
	90	≥ 5	≥ 2	≥ 5	≥ 5

Discusión

El objetivo de este trabajo radicó en el estudio psicométrico del Cuestionario de Adaptación para Adolescentes de Bell (CAAB; Bell, 1934) en población de adolescentes bogotanos. De los análisis aquí realizados se puede concluir que la estructura tetrafactorial propuesta por Bell resultó replicable en una muestra de adolescentes bogotanos. Se realizó la modificación de la consigna y la redacción de algunos ítems para la comprensión de la población a la que va dirigida el cuestionario.

Aunque con un número menor de ítems, se replica la estructura de cuatro factores encontrada en la adaptación española realizada por Cerdá (1987) y la mexicana realizada por Del Bosque-Fuentes y Aragón-Borja (2008). Las cuatro dimensiones evalúan claramente y de manera independiente los cuatro aspectos indicados por Bell: familiar, de salud, social y emocional. Los análisis indicaron que el Cuestionario de Adaptación para Adolescentes de Bell (CAAB) posee cualidades psicométricas adecuadas –evidencias de validez de contenido, aparente y de constructo, así como alta consistencia interna– para realizar una evaluación confiable y válida del constructo adaptación en población de adolescentes bogotanos.

A pesar de que la versión utilizada es mucho más breve que la original, el contenido de cada dimensión logra abordar aquellas dimensiones postuladas originalmente por Bell (1934) y la versión española por Cerdá (1987). Esta versión nueva generada para el contexto colombiano va así en línea con los estándares vigentes internacionalmente en cuanto a la psicometría en los que se pone especial énfasis en contar con versiones más breves para acortar los tiempos de recolección de datos y favorecer la calidad de respuesta en términos de atención y colaboración del evaluado (Moore, Halle, Vandivere, & Mariner, 2002).

Además, a partir de los resultados de las

pruebas *t*, se encontraron diferencias según el sexo en las escalas emocional y familiar para la adaptación adolescente, en las cuales los varones presentaron mayores índices de adaptación que las mujeres. Estos hallazgos son semejantes a los resultados obtenidos en las investigaciones realizadas por Aragón-Borja y Bosques (2012), Funes et al. (2010) y Alonso-Fernández (2005) quienes señalan que las mujeres presentan menores niveles de adaptación en correspondencia con los cambios corporales, el afán por adaptarse al medio que propende autoevaluaciones y la relaciones familiares, las cuales cumplen un papel modelador.

En cuanto a las diferencias por edad refiriendo a las subetapas de la adolescencia, se realizó una prueba ANOVA, que halló diferencias en la escala de adaptación familiar, específicamente entre las subetapas adolescencia nuclear y tardía, y en la escala de adaptación emocional en cada una de las subetapas. Los resultados muestran mayor adaptación para cada dimensión en las subetapas de adolescencia tardía y primaria frente a la adolescencia nuclear, lo que coincide con lo teóricamente expuesto por Aragón-Borja y Bosques (2012), Martínez-Antón, Buelga y Cava (2007), Cardenal-Hernández y Fierro-Bardaji (2001), Carvajal-Corzo (1993) y Flores (1992). Esto está relacionado con los patrones representativos de las actitudes, cambios, vinculaciones y comportamientos de cada una de las subetapas de la adolescencia.

Entre las limitaciones del trabajo se encuentran el hecho de haber utilizado una muestra no probabilística para los análisis y la baja representación de los estratos 1 y 5, lo cual limita la generalización de los resultados a toda la población de adolescentes de Bogotá, Colombia. Sumado a ello, futuras investigaciones deberían valorar la estabilidad de las puntuaciones mediante un test-retest con un intervalo breve de tiempo que

considere las diferencias en la adaptación encontradas entre las subetapas de la adolescencia. Además, sería relevante la continuidad de búsqueda de evidencias psicométricas en términos de validez para verificar, por ejemplo, la invarianza factorial en distintas muestras, así como estudios de validez convergente y discriminante.

El CAAB resulta ser así una herramienta de fácil administración y puntuación fundamentada en la concepción de la adaptación como un comportamiento universal de suma relevancia en la adolescencia (Borges del Rosal, Hernández, & Rodríguez-Naveiras, 2011; Labajos-Alonso, 1996; Rodríguez et al., 1999). Es importante tener en cuenta que el instrumento, además de ser construido sobre este modelo, hace hincapié en la importancia que se da a la adaptación como constructo fundamentado en la autopercepción y la evaluación de sí mismo. Se tiene en cuenta un doble proceso: el ajuste a las preferencias y necesidades propias del adolescente y el ajuste de las conductas al entorno. Se debería llegar entonces a un acuerdo entre las características personales y las demandas del medio, enmarcado en el proceso ambivalente y de constante desarrollo que deviene del período adolescente (Alcalay et al., 2005; Fierro & Cardenal 1996; García-Pérez & Magaz-Lagos, 2001; Gómez-Bustamante & Cogollo, 2010; Martínez-Ferrer et al., 2011).

Se espera que el instrumento aquí presentado –ver Anexo- sea de beneficio para su uso en ámbitos de aplicación y de investigación. Como se ha señalado, el grado de adaptación de los adolescentes no sólo resulta una variable central en este período del ciclo vital, sino que investigaciones precedentes lo han vinculado directamente al desenlace exitoso o fracasado en la transición por este período. El CAAB podría ser utilizado para conocer en qué áreas debe profundizarse la atención a los adolescentes, lo que podría derivar en el diseño de programas de intervención, ya sea en

instituciones educativas o en ámbitos de atención de la salud, con el fin de fomentar las herramientas necesarias en el adolescente para que logre adaptarse a todos los desafíos inherentes a este período de su vida.

Referencias

- Alcalay, L., Milicic, N., & Torretti, A. (2005). Alianza efectiva familia-escuela: Un programa audiovisual para padres. *Psykhé*, 14(2), 149-161. doi: 10.4067/s0718-22282005000200012
- Alba, L. H. (2007). Factores de riesgo para iniciar el consumo de tabaco. *Revista Colombiana de Cancerología*, 4(11), 250-257. Recuperado de <http://www.cancer.gov.co/documentos/revistas/2007/pub4/5.%20Art%-C3%ADculo%20de%20revisi%C3%B3n.pdf>
- Allport, G. W., & Allport, F. H. (1939). *The A-S Reaction Study: A scale for measuring ascendance-submission in personality; manual of directions, scoring values, and norms*. Boston: Mifflin.
- Alonso-Fernández, M. (2005). *Relaciones familiares y ajuste en la adolescencia* (Tesis de doctorado). Universidad de Valladolid, España. Recuperado de https://www.uv.es/lisis/otras-publica/tesis_manuela.pdf
- Alonso-García, J., & Román-Sánchez, J. M. (2005). Prácticas educativas familiares y autoestima. *Psicothema*, 17(1), 76-82. Recuperado de <http://www.psicothema.es>
- Alulema-Dávila, L., & Tintin-Rea, S. (2014). *Adaptación, rasgos de personalidad y factores sociodemográficos comunes en los adolescentes pertenecientes a grupos étnicos católicos salesianos*. (Tesis de grado). Recuperado de <dspace.uazuay.edu.ec/bitstream/datos/3656/1/10333.pdf>
- Aragón-Borja, L. E., & Bosques, E. (2012). Adaptación familiar, escolar y personal de adolescentes en la ciudad de México. *Enseñanza e Investigación en Psicología*, 17(2), 263-282. Recuperado de <http://www.redalyc.org/articulo.oa?id=29224159002>

- Atienza-González, F. L., Moreno-Sigüenza, Y., & Balaguer-Solá, I. (2000). Análisis de la dimensionalidad de la Escala de Autoestima de Rosenberg en una muestra de adolescentes valencianos. *Revista de Psicología Universitas Tarraconensis*, 22(2), 29-42.
- Ayala-Velázquez, H., Pedroza-Cabrera, F., Morales-Chainé, S., Chaparro Caso-López, A., & Barragán-Torres, N. (2002). Factores de riesgo, factores protectores y generalización del comportamiento agresivo en una muestra de niños en edad escolar. *Salud Mental*, 25(3), 27-40. Recuperado de <http://www.redalyc.org/pdf/582/58232504.pdf>
- Baessler, J., & Schwarzer, R. (1996). Evaluación de la autoeficacia: Adaptación española de la Escala de Autoeficacia General. *Ansiedad y Estrés*, 2(1), 1-8.
- Bell, H. M. (1934). *The Adjustment Inventory*. California: Stanford University Press.
- Bernreuter, R. (1935). *Manual for the Personality Inventory*. California: Stanford University Press.
- Bharvad, M. B. (2015). A study of emotional intelligence and adjustment among school students. *The International Journal of Indian Psychology*, 2(2), 23-31. Recuperado de <http://oaji.net/articles/2015/1170-1427635156.pdf>
- Borges del Rosal, A., Hernández, C., & Rodríguez-Naveiras, E. (2011). Evidencias contra el mito de la inadaptación de las personas con altas capacidades intelectuales. *Psicothema*, 23(3), 362-367. Recuperado de <http://www.psicothema.com/>
- Caballo, V. E., & Ortega, A. R. (1989). La escala multidimensional de Expresión Social: Algunas propiedades psicométricas. *Revista de Psicología General y Aplicada*, 42(2), 215-221. Recuperado de <https://dialnet.unirioja.es/servlet/articulo?codigo=2359241>
- Cardenal-Hernández, V., & Fierro-Bardaji, A. (2001). Sexo y edad en estilos de personalidad, bienestar social y adaptación social. *Psicothema*, 13(1), 118-126. Recuperado de <http://www.psicothema.com/>
- Carvajal-Corzo, G. (1993). *Adolecer: La aventura de una metamorfosis. Una visión psicoanalítica de la adolescencia*. Bogotá: Printing Service Network.
- Cerdá, E. (1987). *Cuestionario de Adaptación para Adolescentes. Manual*. Barcelona, España: Herder.
- Coleman, J., & Hendry, L. (1999). *The nature of adolescence* (3ª ed.). Barcelona, España: Routledge.
- Del Bosque-Fuentes, A. E., & Aragón-Borja, L. E. (2008). Nivel de adaptación en adolescentes mexicanos. *Interamerican Journal of Psychology*, 42(2), 287-297. Recuperado de <http://pepsic.bvsalud.org/pdf/rip/v42n2/v42n2a10.pdf>
- Departamento Administrativo Nacional de Estadística (DANE). (2005). *Análisis de la estructura y composición de las principales variables demográficas y socioeconómicas del Censo 2005*. Recuperado de https://www.dane.gov.co/files/censos/Grupo_mixto22%20_PUBL.pdf
- 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
- Fernández-Ballesteros, R., & Sierra, B. (1982). Estudio factorial sobre la percepción del ambiente escolar. En R. Fernández-Ballesteros (Ed.), *Evaluación de Contextos* (pp. 9-49). Murcia: Servicio de Publicaciones de la Universidad de Murcia.
- Fierro, A. (1990). Autoestima en adolescentes. Estudios sobre su estabilidad y sus determinantes. *Estudios de Psicología*, 12(45), 85-107. doi: 10.1080/02109395.1991.10821156
- Fierro, A. (1997). Estrés, afrontamiento y adaptación. En M. I. Hombrados (Ed.), *Estrés y salud* (pp. 11-37). Valencia, España: Promolibro.
- Fierro, A., & Cardenal, V. (1996). Dimensiones de personalidad y satisfacción personal. *Revista de Psicología General y Aplicada*. 49(1), 65-82. Recuperado de <https://dialnet.unirioja.es/servlet/articulo?codigo=2358113>
- Flores, A. (1992). Educación sexual. Uruguay: Dismar.
- Funes, M. J., Lupiáñez, J., & Humphreys, G. (2010). Analyzing the generality of conflict adaptation effects. *Journal of Experimental Psychology*, 36(1), 147-161. doi: 10.1037/a0017598

- García-Pérez, E. M., & Magaz-Lagos, A. (2001). *Escala Magallanes de Adaptación*. Madrid, España: Albor COHS.
- Gómez-Restrepo, C. (2005). Psiquiatría y salud mental de niños y adolescentes: Una necesidad. *Revista Colombiana de Psiquiatría*, 34(3), 339- 345. Recuperado de http://www.scielo.org.co/scielo.php?pid=S0034-74502005000300001&script=sci_arttext&tlng=pt
- Gómez-Bustamante, E. M., & Cogollo, Z. (2010). Factores predictores relacionados con el bienestar general en adolescentes estudiantes de Cartagena, Colombia. *Revista de Salud Pública*, 12(1), 61-70. doi: 10.1590/s0124-00642010000100006
- González, A. L., De los Ríos, D. P., & Viveros, E. F. (2015). Crisis familiares y rendimiento académico en niños de 8 a 10 años. Una aproximación desde la dinámica interna familiar. *Infancias Imágenes*, 14(2), 25-36. doi: 10.14483/udistrital.jour.infigm.2015.2.a02
- Gutiérrez, G. (2000). *Adaptación, validación y estandarización de la Escala Multidimensional de Expresión Social –Parte Motora y Parte Cognitiva en estudiantes de la Universidad Nacional de Colombia* (Tesis de grado). Bogotá, Colombia: Universidad Nacional de Colombia.
- Holahan, C. J., Moos, R. H., & Schaefer, J. A. (1996). Coping, stress resistance, and growth: Conceptualizing adaptive functioning. En M. Zeidner & N. S. Endler (Eds.), *Handbook of coping: Theory, Research and Application* (pp. 24-42). New York: John Wiley & Sons.
- Huebner, E. S. (1991). Initial development of the Student's Life Satisfaction Scale. *School Psychology International*, 12(3), 231-240. doi: 10.1177/0143034391123010
- Jiménez, M. I., & López-Zafra, E. (2011). Actitudes sociales y adaptación social en adolescentes españoles: El papel de la inteligencia emocional percibida. *Revista de Psicología Social*, 26(1), 105-117. doi: 10.1174/021347411794078417
- Kostanski, M., & Gullone, E. (1998). Adolescent body image dissatisfaction: Relationship with self-esteem, anxiety, and depression controlling for body mass. *Journal of Child Psychology and Psychiatry*, 39(2), 255-262. doi: 10.1017/s0021963097001807
- Kovacs, M. (2004). *Inventario de Depresión Infantil (CDI)*. Madrid, España: TEA Ediciones.
- Kulshrestha, S. P. (1979). *Educational Psychology*. Meerut, India: Loyal Book Depot.
- Labajos-Alonso, J. (1996). Identidad del adolescente. En Á. Aguirre (Ed.), *Psicología de la Adolescencia* (pp. 173-194). Bogotá, Colombia: Alfaomega.
- Ley 142. (1994). *Ley Sobre Servicios Públicos*. Recuperado de <https://www.minminas.gov.co/documents/10180/670382/LEY142DE1994.pdf/68f0c21d-fd78-4242-b812-a6ce94730bfl>
- Lévy-Mangin, J. P., & Varela-Mallou, J. (2006). *Modelización con estructuras de covarianzas en ciencias sociales*. Coruña, España: Netbiblio.
- Lynch, D., Norem, K., & Gergen, R. (1981). *Self-Concept. Advances in theory and research* (pp. 119-13). Cambridge: Ballinger Publish.
- Macías, V. T. (2000). *Ser adolescente*. México: Trillas.
- Martínez-Arias, M. R. (1995). *Psicometría: Teoría de los test psicológicos y educativos*. Madrid, España: Síntesis.
- Martínez-Antón, M., Buelga, S., & Cava, M. J. (2007). La satisfacción con la vida en la adolescencia y su relación con la autoestima y el ajuste escolar. *Anuario de Psicología*, 38(2), 293-303. Recuperado de <http://www.redalyc.org/articulo.oa?id=97017404013>
- Martínez-Ferrer, B., Amador-Muñoz, L.V., Moreno-Ruiz, D., & Musitu-Ochoa, G. (2011). Implicación y participación comunitaria y ajuste psicosocial en adolescentes. *Psicología y Salud*, 21(2), 205-214. Recuperado de <https://www.uv.es/lisis/belen/12-psicologiasalud.pdf>
- Minda-Mina, J. J. (2011). *Efectos emocionales presentes en adolescentes de doce a quince años de edad, procedentes de hogares disfuncionales en los que cuentan con solo la madre o el padre y su influencia en el*

- comportamiento y adaptación, en el Instituto Tecnológico Superior Los Shyris, Año lectivo 2010 - 2011.* (Tesis de grado). Recuperado de www.dspace.uce.edu.ec/bitstream/25000/1399/1/T-UCE-0007-6.pdf
- Moos, R. H. (1974). *The Social Climate Scales: An overview*. Palo Alto, California: Consulting Psychologists Press.
- Moraleda, M., González-Galán, A., & García-Gallo, J. (1998). *AECS. Actitudes y estrategias cognitivas sociales*. Madrid, España: TEA.
- Moore, K. A., Halle, T. G., Vandivere, S., & Mariner, C. L. (2002). Scaling back survey scales. How short is too short? *Sociological Methods & Research*, 30(4), 530-567. doi: 10.1177/0049124102030004003
- Musitu, G., & García, F. (2001). *ESPA29: Escala de Estilos de Socialización Parental en la Adolescencia*. Madrid, España: TEA
- Ochoa, M. A. (2012). *Adaptación de adolescentes en riesgo de delincuencia*. (Tesis de grado). Universidad Rafael Landívar. Guatemala.
- Oliva-Delgado, A., & Parra-Jiménez, A. (2004). Contexto familiar y desarrollo psicológico durante la adolescencia. En E. Arranz (Ed.), *Familia y desarrollo psicológico* (pp. 96-123). Madrid, España: Pearson Prentice-Hall.
- Orte-Socias, C., & March-Cerdá, M. (1996). *Pedagogía de la Inadaptación Social: Educación Social*. Valencia, España: Nau Libres.
- Palmero, F., & Fernández E. (1998). *Emociones y adaptación*. Barcelona, España: Ariel.
- Petersen, A. C., & Crockett, L. (1985). Pubertal timing and grade effects on adjustment. *Journal of Youth and Adolescence*, 14(3), 191-206. doi: 10.1007/bf02090318
- Profamilia (2016). Informe Anual de Actividades. Recuperado de <http://www.profamilia.org.co/docs/INFORME%20PROFAMILIA%202016%20VERSION%20MARZO%202016.pdf>
- Programa Rumbos (2001). Encuesta nacional sobre consumo de sustancias psicoactivas en jóvenes de 10-24 años. Recuperado de https://www.unodc.org/documents/colombia/Documentostecnicos/Estudio_Consumo_Escolares.pdf
- Rabazo-Méndez, M. (1999). *Interacción Familiar, Competencia socio-escolar y comportamiento disocial en adolescentes* (Tesis Doctoral). Universidad de Extremadura, Facultad de Educación. Departamento de Psicología y Sociología de la Educación. Recuperado de <https://dialnet.unirioja.es/descarga/tesis/337.pdf>
- Rodríguez-Sabiote, C., Gallardo-Vigil, M. A., Pozo-Lorente, T., & Gutiérrez-Pérez, J. (2006). *Iniciación al análisis de datos cuantitativos en educación. Análisis descriptivo básico: teoría y práctica mediante SPSS*. Recuperado de <http://www.ugr.es/~erivera/PaginaDocencia/Posgrado/Documentos/ClementeCuadernoDescriptiva.pdf>
- Rodríguez, M. C., Jiménez, M. A., Fernández, E., & Godoy, C. (1999). Evaluación de psicopatología en la infancia y adolescencia a través de la percepción de los padres. *Revista Iberoamericana de Diagnóstico y Evaluación Psicológica*, 8(2), 51-65. Recuperado de http://www.aidep.org/03_ridep/R08/R083.pdf
- Rodríguez-Mayoral, J. M., Martínez-Arias, R., Díaz-Aguado, M. J., & Morentín, R. (2008). Comportamiento Violento en Adolescentes: Su relación con las estrategias cognitivas y el rendimiento académico. *Psicología Educativa*, 14(1), 63-81.
- Ros-Rahola, R., Morandi-Garde, T., Cozzetti-Sueldo, E., Lewintal-Blaustein, C., Cornella i Canals, J., & Surís-Granell, J.C., (2001). La adolescencia: Consideraciones biológicas, psicológicas y sociales. En C. Buil-Rada, I. Lete-Lasa, R. Ros-Rahola & J. L. De Pablo-Lozano (Eds.), *Manual de Salud Reproductiva en la Adolescencia* (pp. 27-84). Wyeth-Lederle: España.
- Rosenberg, M. (1966). *Society and the adolescent self-image*. Princeton: University Press.
- Siverio-Eusebio, M. A., & García-Hernández, M. D. (2007). Self-perception of adjustment and sadness in adolescence. *Anales de Psicología*, 23(1), 41-48. Recuperado de http://www.um.es/analesps/v24/v24_1i.htm

- Suárez, E., & Krauskopf, D. (1995). El enfoque de riesgo y su aplicación a las conductas del adolescente. Una perspectiva psicosocial. En M. Maddaleno, M. Munist, C. Serrano, T. Silver, E. Suárez & J. Yunes (Eds.), *La salud del Adolescente y el Joven*. Washington, O.P.S. Publicación científica n° 552, pp. 183-193.
- Thurstone, L.L. (1930). *Instructions for using the Personality Schedule*. Chicago, Illinois: Univ. of Chicago Press.
- Valverde, J. (1988). *El proceso de inadaptación social*. Madrid, España: Popular.
- Villarduna-Ríos, M. M. (2013). Adaptación del clima social familiar en padres de familia de instituciones educativas nacionales del distrito de San Martín. *Minds*, 1(1), 45-65. <http://ojs.ucvlima.edu.pe/index.php/minds/article/view/15>
- Vinaccia, S., Quiceno, J. M., & Moreno-San-Pedro, E. (2007). Resiliencia en Adolescentes. *Revista Colombiana de Psicología*, 16, 139-146. Recuperado de <http://www.redalyc.org/pdf/804/80401610.pdf>
-

Anexo 1

Cuestionario de Adaptación para adolescentes BELL - CAAB

Te presentamos una serie de preguntas que tienen como fin conocer tu percepción acerca de ti mismo y del ambiente en general. Te pedimos que respondas las preguntas con toda sinceridad, teniendo en cuenta que no hay respuestas buenas ni malas, por lo que debes marcar tu respuesta según tu realidad y no la que quisieras que fuera.

Marca en la casilla correspondiente a "SI", "NO" o "?", según sea tu respuesta. Debes responder siempre que te sea posible "SI" o "NO" y ÚNICAMENTE marcarás el signo de interrogación "?" en aquellos casos en que estés completamente seguro de no poder responder ni "SI" ni "NO". Si no vives en casa de tus padres, las preguntas que hagan referencia a tu hogar contéstalas con relación a las personas con las que convivas.

Ítem	Afirmación	SI	NO	?
1	Cuándo estás en una reunión o con un grupo de amigos, ¿te gusta ser el que presenta unos a otros?			
2	¿A veces has tomado la iniciativa para animar alguna reunión aburrida?			
3	¿Crees que en tu casa debería existir un poco más de comprensión y de afecto?			
4	¿Sufres con frecuencia heridas o lesiones importantes en tu cuerpo?			
5	¿Lloras con facilidad?			
6	¿Has tenido alguna operación quirúrgica importante?			
7	¿Te resulta difícil entablar conversación con una persona que acabas de conocer?			
8	¿Has tenido dos o más enfermedades importantes durante los últimos 10 años? (enfermedades graves o que te hayan obligado a guardar cama durante varias semanas)			
9	¿Estás casi siempre en desacuerdo con la forma en que tus padres tratan los asuntos de casa?			
10	¿Te desanimas fácilmente?			
11	¿Te parece que tus padres han sido o son demasiado severos contigo?			
12	¿Has necesitado visitar al médico con bastante frecuencia?			
13	¿Te resulta muy difícil hablar en público?			
14	¿Alguno de tus padres tiene costumbres o modales que te desagradan?			
15	¿De vez en cuando te ves obligado a callarte o a marcharte a otra habitación con la finalidad de que haya paz y tranquilidad en tu hogar?			
16	¿Haces amigos con facilidad?			
17	¿Sientes a menudo nerviosismo?			
18	Luego de haber sufrido una situación humillante, ¿te quedas muy preocupado durante bastante tiempo?			
19	¿Faltas a clase con cierta frecuencia por estar enfermo?			
20	¿Te gusta asistir a fiestas y reuniones a las que acuden muchas personas?			
21	¿Tienes altibajos de humor sin alguna razón aparente?			
22	¿Te sientes seguro de ti mismo cuando debes hablar en clase?			
23	A la hora de dormir, ¿Tienes a menudo muchas ideas en tu cabeza a las que le das tantas vueltas que te impiden dormir normalmente?			
24	¿Frecuentemente tienes la impresión de que tus padres no te comprenden bien?			

Clave de corrección

Sumar los ítems de cada dimensión de acuerdo a la siguiente codificación: NO = 0, ? = 1, SI = 1. Previo a ello deben invertirse las puntuaciones de los ítems que tienen la letra R (NO = 1, ? = 1, SI = 0).

Dimensión del CAAB	Número de Ítem
Factor 1. Familiar	3, 9, 11, 14, 15, 24
Factor 2. Salud	4, 6, 8, 12, 19
Factor 3. Social	1R, 2R, 7, 13, 16R, 20R, 22R
Factor 4. Emocional	5, 10, 17, 18, 21, 23

Propiedades Psicométricas del Inventario de Crecimiento Postraumático en Población Ecuatoriana

Patricio R. Arias * ¹, Felipe E. García ², Isabel Patricia Valdivieso ³

1 - Universidad Tecnológica Indoamérica, Ambato, Ecuador.

2 - Facultad de Ciencias Sociales, Universidad Santo Tomás, Chile.

3 - Facultad de Ciencias Humanísticas y Sociales, Universidad Técnica de Manabí, Portoviejo, Ecuador.

Introducción
Metodología
Resultados
Discusión
Referencias

Recibido: 11/07/2017 Revisado: 01/08/2017 Aceptado: 15/08/2017

Resumen

Dentro del estudio del comportamiento humano ante la exposición a un evento altamente estresante, la psicología ha dado un salto al estudiar no sólo sus efectos negativos sino también el aprendizaje que la persona puede extraer de dichas vivencias, lo que se denomina *crecimiento postraumático*. El objetivo de este estudio fue analizar las propiedades psicométricas del Inventario de Crecimiento Postraumático (PTGI) en una muestra de 450 adultos sobrevivientes al terremoto ocurrido en el año 2016 en las costas de Ecuador. El PTGI reveló una alta consistencia interna ($\alpha = .95$). El análisis factorial confirmatorio mostró un ajuste apropiado para una estructura bifactor de tres factores específicos y un factor general. El PTGI además mostró invarianza factorial al comparar hombres y mujeres. Se concluye que el PTGI es un instrumento confiable y con validez interna para investigar el crecimiento postraumático en población ecuatoriana.

Palabras Clave: *psicología positiva, análisis factorial confirmatorio, desastre natural, terremoto*

Abstract

In the study of human behavior concerning the exposure to highly stressful events, psychology has taken a leap forward, studying not only the negative effects of such events, but also the knowledge a person may gain from them, which is known as *posttraumatic growth*. The aim of this study was to analyze the psychometric properties of the Posttraumatic Growth Inventory (PTGI) in a sample of 450 adult survivors of the earthquake that took place on the Ecuadorian coast in 2016. The PTGI showed high internal consistency ($\alpha = .95$). The confirmatory factor analysis displayed an appropriate adjustment for a bifactor structure consisting of three specific factors and one general factor. Furthermore, the PTGI showed a factorial invariance when comparing men and women. We could conclude that the PTGI is a reliable and internally valid instrument for the study of posttraumatic growth in Ecuadorian population.

Keywords: *positive psychology, confirmatory factorial analysis, natural disaster, earthquake*

*Correspondencia a: Patricio Arias, info@patricioarias.com

Cómo citar este artículo: Arias, P. R., García, F. E., & Valdivieso, I. P. (2017). Propiedades psicométricas del Inventario de Crecimiento Postraumático en población ecuatoriana. *Revista Evaluar*, 17(2), 67-80. Recuperado de <https://revistas.unc.edu.ar/index.php/revaluar>

Introducción

Un sobreviviente a un evento en el que ha corrido riesgo su vida, la de personas cercanas o que ha presenciado este riesgo en otras personas puede llegar a sufrir síntomas de estrés postraumático ([American Psychiatric Association, 2013](#)); sin embargo, muchos de ellos son capaces de percibir cambios positivos luego de estos eventos ([Tedeschi & Calhoun, 1996, 2000](#)), lo que les permite alcanzar un mayor bienestar y ajuste psicológico y controlar a largo plazo la aparición de trastornos mentales ([García, Cova-Solar, & Melipillán, 2013](#)). A este fenómeno se lo ha denominado *crecimiento postraumático* (CPT). El objetivo principal del presente estudio fue evaluar las propiedades psicométricas de un instrumento que permita medir estos cambios en población ecuatoriana afectada por un desastre natural.

El CPT abarca tres dimensiones conceptuales: a) *cambios en la autopercepción*, que incluyen sentirse más fuerte, seguro y capaz de afrontar dificultades en el futuro; b) *cambios en la relación con otros*, que incluyen mayor cercanía con la familia y amistades y mayor apertura para recibir ayuda, y c) *cambios en la filosofía de vida*, que se define como una mayor valoración de la propia existencia e implica tomar la vida con más calma y disfrutar el momento ([Tedeschi & Calhoun, 2008](#); [Calhoun, Cann, & Tedeschi, 2010](#)). Estas tres dimensiones teóricas no se han traducido necesariamente en una cantidad equivalente de factores cuando se ha evaluado el instrumento diseñado para medir estos cambios: el Inventario de Crecimiento Postraumático (PTGI, por sus siglas en inglés).

El PTGI es un instrumento que consta de 21 ítems y que mostró en su versión original poseer una alta consistencia interna ($\alpha = .90$) y estabilidad temporal ($r = .71$). A través de un análisis factorial exploratorio se determinó que estaba cons-

tituido por cinco factores ([Tedeschi & Calhoun, 1996](#)): 1) *relación con otros*, incluye ítems acerca del fortalecimiento de las relaciones interpersonales, (2) *apreciación de la vida*, incluye ítems acerca del aumento del valor que la persona da a la vida, (3) *fuerza personal*, incluye ítems acerca de una mejor percepción de las propias capacidades, habilidades y resistencia personal, (4) *cambio en la filosofía de la vida*, incluye ítems sobre la reevaluación del sentido y significado de la vida tras la experiencia traumática y (5) *nuevas posibilidades*, incluye ítems acerca del aumento en la sensación de que la vida brinda nuevas oportunidades y prioridades.

El PTGI se ha utilizado para trabajar con poblaciones que han vivido diversos eventos traumáticos tales como desastres naturales ([García, Jaramillo, Martínez, Valenzuela, & Cova-Solar, 2014](#)), violencia ([Hooper, Marotta, & Depuy, 2009](#)), diagnósticos de enfermedades que amenazan la vida ([Brunet, McDonough, Hadd, Crocker, & Sabiston, 2010](#)), guerras ([Palmer, Graca, & Occhietti, 2012](#)) o eventos diversos ([García, Páez, Reyes-Reyes, & Álvarez, 2017](#)). Este instrumento ha llegado a ser el más utilizado para el estudio de los efectos positivos de una vivencia potencialmente traumática ([Joseph & Linley, 2008](#)).

Las adaptaciones realizadas al PTGI no siempre han mostrado la estructura interna de los cinco factores que encontraron [Tedeschi y Calhoun \(1996\)](#). Además, se han encontrado estructuras que varían desde uno a cuatro factores, entre ellas la estructura de tres dimensiones que propone el modelo teórico original, como se puede observar en la Tabla 1.

En el análisis de las propiedades psicométricas del PTGI de [Joseph, Linley y Harris \(2004\)](#) se consideró un modelo de un solo factor por la fuerte relación entre sí que habían mostrado los cinco factores iniciales, sin embargo los autores aceptaron la posibilidad de que el instrumento

Tabla 1
Estructuras factoriales encontradas en el PTGI.

Estructura factorial	Autor	País	Muestra
Un factor	Joseph et al. (2004)	Reino Unido	176 pacientes que han experimentado acontecimientos angustiosos de la vida
	Costa-Requena & Gil-Moncayo (2007)	España	130 pacientes con diagnóstico de cáncer
Tres factores correlacionados	Powell et al. (2003)	Sarajevo, Bosnia y Herzegovina	136 adultos refugiados y desplazados expuestos a diversos eventos traumáticos
	Anderson & Lopez-Baez (2008)	EE.UU.	345 estudiantes universitarios
Modelo jerárquico de tres factores de primer orden y uno de segundo orden	García et al. (2013)	Chile	449 sobrevivientes a un desastre natural
Modelo bifactor de cinco factores específicos y un factor general	Konkolý-Thege et al. (2014).	Hungría	691 personas que experimentaron algún tipo de trauma o pérdida
Modelo bifactor de tres factores específicos y un factor general	Rodríguez-Rey et al. (2016)	España	143 padres de niños hospitalizados
	Garrido-Hernansaiz et al (2017)	España	304 adultos seropositivos
Modelo de cinco factores correlacionados	Brunet et al. (2010)	Canadá	470 pacientes con cáncer
	Esparza-Baigorria et al. (2016)	Argentina	156 pacientes oncológicos
	Hooper et al. (2009)	EE.UU.	143 estudiantes parentalizados
	Palmer et al. (2012)	EE.UU	221 veteranos de guerra
	García da Silva et al. (2016)	Brasil	300 estudiantes universitarios
	Taku et al. (2008)	Varios	926 personas expuestas a distintos eventos traumáticos

incluyera tres factores. De la misma manera, utilizando un análisis exploratorio de componentes principales (ACP) se encontraron tres factores tanto en estudiantes universitarios (Anderson & López-Báez, 2008) como en sobrevivientes de guerra (Powell, Rosner, Butollo, Tedeschi, & Calhoun, 2003); estos dos estudios conservaron los factores de *cambio en la relación con otros*

y de *filosofía de la vida*, pero *nuevas posibilidades*, *apreciación de la vida* y *fuerza personal* se reunieron en un solo factor al que llamaron *cambios en la autopercepción*. Con el análisis factorial confirmatorio del PTGI traducido al español se obtuvo en Chile esa misma estructura de tres factores de primer orden y un factor de segundo orden (García et al., 2013) en un estudio con per-

sonas adultas que fueron afectadas por el terremoto del año 2010. A su vez, en Argentina se ha encontrado la estructura original de cinco factores correlacionados (Esparza-Baigorria, Leibovich de Figueroa, & Martínez-Terrer, 2016) en mujeres adultas afectadas por cáncer de mama. Por otra parte, Rodríguez-Rey, Alonso-Tapia, Kassam-Adams y Garrido-Hernansaiz (2016) y Garrido-Hernansaiz, Rodríguez-Rey y Alonso-Tapia (2017) obtuvieron en España, con una versión de 12 y 11 ítems respectivamente, un modelo bifactor con tres factores específicos: *cambios en la filosofía de la vida*, *cambio en la autopercepción* y *cambios en la relación con otros*, y un factor general, utilizando la traducción para población hispana de los EE.UU (Weiss & Berger, 2006).

En el estudio con diversas poblaciones que utilizó el PTGI ha llamado la atención la diferencia de género, de modo que las mujeres informan mayor CPT que los hombres. Un metaanálisis que incluyó 70 estudios llevó a la conclusión de que en el CPT existen diferencias de género modestas, pero significativas entre hombres y mujeres, siendo las mujeres las que muestran más crecimiento que los hombres (Vishnevsky, Cann, Calhoun, Tedeschi, & Demakis, 2010). Esta diferencia se ha observado también en población latinoamericana (García, Capponi, et al., 2016; García, Jaramillo, et al., 2014).

En el caso de Ecuador, al ser un país expuesto a constantes riesgos de origen natural, la adaptación del PTGI puede contribuir al estudio de las respuestas psicológicas frente a las amenazas, lo que permitirá estudiar los cambios positivos que los ecuatorianos sobrevivientes al terremoto de abril de 2016 han presentado tras el sismo. De ese modo, el objetivo del presente estudio es evaluar las propiedades psicométricas de una traducción del PTGI hecha para población ecuatoriana, para lo cual se evaluará su estructura factorial y la invarianza factorial entre grupos de hombres y mu-

jes, además de los estadísticos descriptivos y la consistencia interna del instrumento. Hasta hoy, no se han realizado estudios que hayan utilizado esta escala en Ecuador, por lo que este sería el primer estudio que evalúe sus propiedades psicométricas en esta población.

Metodología

Diseño

El diseño utilizado tiene un carácter no experimental, es de naturaleza descriptiva y transversal y es de índole instrumental.

Participantes

Participaron 450 individuos, 256 mujeres (56.9%) y 194 hombres (43.1%). La edad promedio fue de 26.7 (DE = 10.66) con un rango de 18 a 65 años. Los participantes son pobladores sobrevivientes al terremoto del 16 de abril del año 2016 en la provincia de Manabí.

Instrumentos

Se utilizó el Inventario de Crecimiento Postraumático (PTGI; Tedeschi & Calhoun, 1996), el cual consta de 21 ítems que se responden en una escala Likert desde 0 (*Ningún cambio*) hasta 5 (*Un cambio enorme*). Se realizó una traducción y traducción inversa para obtener una versión adecuada para la población ecuatoriana (ver Anexo). Las propiedades psicométricas del instrumento en el presente estudio se detallan en la sección de Resultados. Además se incluyó un cuestionario socio-demográfico breve en el que se registró el sexo, la edad y el nivel de estudio de los participantes.

Procedimiento

En primer lugar se realizó una nueva traducción del instrumento distinta a las ya disponibles con la finalidad de evitar modismos lingüísticos de otros países, considerando que la conformación multiétnica y multicultural ecuatoriana podría implicar construcciones lingüísticas diferentes. Para ello, un experto psicólogo que maneja los dos idiomas realizó una traducción directa de la escala original del inglés al español, luego otro psicólogo experto realizó una traducción inversa para comparar las coincidencias lingüísticas. El resultado final se sometió a juicio experto con un investigador que ha estudiado el CPT en otra población latinoamericana y se aplicó en una prueba piloto con 70 estudiantes universitarios que fueron afectados por el terremoto, quienes verificaron la comprensión de las instrucciones y de los ítems. Un análisis de consistencia interna arrojó buenos resultados ($\alpha = .95$ para la escala total).

Para el cálculo del tamaño muestral adecuado para la población objetivo se siguió el criterio recomendado por [Martínez-Arias \(1995\)](#), quien sugiere un mínimo de 300 participantes o entre 5 y 10 por ítem. Siguiendo este último criterio, y dado que el PTGI consta de 21 ítems, la muestra óptima debía incluir un mínimo de 210 participantes. Finalmente se encuestó a 450 personas, lo que permitió que los análisis presentados tuvieran suficiente potencia.

El estudio se realizó en la población de la provincia de Manabí sobreviviente al terremoto del 16 de abril de 2016, para lo cual se solicitó la colaboración de docentes de una universidad de la zona, a quienes se capacitó para la aplicación del instrumento. Ellos a su vez solicitaron a sus estudiantes que lo aplicaran a personas de su contexto social cercano.

Antes de responder al instrumento, se pidió a los participantes firmar un formulario de con-

sentimiento informado en el que se aseguraba el anonimato, la confidencialidad de los datos, la voluntariedad para responder y el resguardo de los datos por parte de los investigadores.

Análisis de datos

Se realizó un análisis factorial confirmatorio (AFC) poniendo a prueba cinco modelos: el primero de ellos fue el modelo de un factor ([Joseph et al., 2004](#)); el segundo fue el modelo de tres factores jerárquico ([García et al., 2013](#)) que, debido a que se encontró en población latinoamericana también afectada por un terremoto, constituyó el modelo principal evaluado en este estudio; el tercero fue el modelo bifactor de tres factores específicos y un factor general, encontrado en recientes estudios con población española ([Garrido-Hernansaiz et al., 2017](#); [Rodríguez-Rey et al., 2016](#)); el cuarto fue el modelo de cinco factores jerárquico, equivalente al obtenido en Argentina ([Esparza-Baigorria et al., 2016](#)) y finalmente se evaluó un modelo bifactor de cinco factores específicos y un factor general, encontrado en una muestra húngara ([Konkolý-Thege, Kovács, & Balog, 2014](#)).

Para los análisis se utilizó el método de estimación de máxima verosimilitud, previa evaluación de la normalidad multivariante a través del coeficiente de [Mardia \(1970, 1974\)](#); existe normalidad cuando los valores se ubican dentro del rango ± 5 ([Bentler, 2005](#)) pero se considera adecuado utilizar el método de estimación de máxima verosimilitud cuando los valores no exceden el rango ± 70 ([Rodríguez-Ayán & Ruiz-Díaz, 2008](#)). La calidad del ajuste de los modelos analizados se evaluó en base a los siguientes estadísticos y criterios de buen ajuste ([Byrne, 2006](#); [Kline, 2005](#)): a) χ^2 : un valor no significativo se considera un ajuste adecuado; b) χ^2/df : un valor menor a 3 indi-

Tabla 2
Índices de ajuste de los modelos.

Modelo	χ^2 (gl)	χ^2 (gl)	CFI	TLI	AIC	SRMR	RMSEA (90% IC)
1: Un factor	805.02*** (189)	4.26	.90	.89	931.02	.04	.09 (.08/.09)
2: Tres factores-jerárquico	579.17*** (186)	3.11	.93	.94	711.17	.04	.07 (.06/.08)
3. Tres factores-bifactor	365.83*** (165)	2.22	.97	.96	497.83	.03	.05 (.05/.06)
4: Cinco factores-jerárquico	594.07*** (184)	3.18	.93	.94	730.07	.04	.07 (.06/.08)
5. Cinco factores-bifactor	351.58*** (158)	2.23	.97	.96	497.58	.03	.05 (.05/.06)

Nota. *** $p < .001$

ca buen ajuste y aceptable cuando es menor de 5; c) CFI y TLI: un valor .95 o superior es indicador de un ajuste adecuado, un valor sobre .90 indica un ajuste aceptable; d) AIC: es un índice comparativo, por lo que el modelo que presente el valor más bajo es el que mejor se ajusta a los datos; e) SRMR: un valor inferior a .08 es un ajuste aceptable; f) RMSEA e intervalo de confianza del 90%: un valor inferior a .08 es un ajuste aceptable, el intervalo de confianza debe tener menos que un .10 en su valor superior.

Se realizó una comparación de grupos a través de *t* de Student para ver si existían diferencias por sexo. Luego se estudió la invarianza factorial entre los participantes hombres y mujeres a través de un análisis multi-muestra. Para ello, se realizó la evaluación secuencial de la invarianza configural, métrica, fuerte y estricta (Elosúa, 2005). La invarianza configural implica que los factores están especificados por los mismos ítems en los dos grupos. La invarianza métrica evalúa la igualdad de coeficientes de regresión. La invarianza fuerte evalúa la igualdad en los interceptos. La invarianza estricta evalúa la igualdad en los errores y representa el mayor nivel de acuerdo alcanzable entre estructuras factoriales.

Tradicionalmente, la prueba de invarianza entre niveles se realiza estudiando las diferen-

cias en el estadístico χ^2 entre un nivel y otro, ya que los modelos se encuentran anidados (Bollen, 1989). Como este estadístico es sensible al tamaño de la muestra, es probable que se rechace la invarianza en muestras grandes, por lo que se consideró principal indicio de invarianza que el CFI no varíe más de .01 con respecto al modelo anterior (Cheung & Rensvold, 2002). Se espera además un TLI superior a .90 y un RMSEA inferior a .08.

Se empleó el estadístico alfa de Cronbach para estimar la consistencia interna del PTGI total y de sus factores, así como la correlación ítem-total corregida. Por último, se realizó un análisis descriptivo del instrumento y sus factores, así como una evaluación de la correlación bivariada entre ellos. Los análisis se llevaron a cabo empleando el programa SPSS v.21 (IBM Corp., 2011) y AMOS SPSS v.20. (Arbuckle, 2011)

Resultados

Para realizar el AFC se evaluó inicialmente el coeficiente de Mardia, que arrojó un valor de 43.09, considerado adecuado para el uso del método de estimación de máxima verosimilitud, por lo que se continuó el análisis con dicho método.

En la Tabla 2 se presentan los resultados ob-

Tabla 3

Estadísticos descriptivos, consistencia interna, correlaciones en PTGI total y sus tres factores.

ESCALA	Mujeres (n = 256)		Hombres (n = 194)		Total (N = 450)		α	Correlaciones			
	M	DE	M	DE	M	DE		1	2	3	4
1. CPT-Autopercepción	37.72	13.72	35.66	14.81	36.84	14.22	0.94	na	.84***	.70***	.97***
2. CPT-Relación con otros	22.98	7.93	20.69	8.92	21.99	8.44	0.91		na	.72***	.93***
3. CPT-Filosofía de vida	6.00	3.04	5.09	3.11	5.61	3.10	0.83			na	.79***
4. CPT Total	66.70	23.03	61.44	25.31	64.44	24.15	0.96				na

Nota. *** $p < .001$; na = no aplica.

tenidos en el AFC en los cinco modelos: a) modelo unifactorial; b) modelo jerárquico de tres factores de primer orden y un factor de segundo orden; c) modelo bifactor de tres factores específicos y un factor general; d) modelo jerárquico de cinco factores de primer orden y un factor de segundo orden, y e) modelo bifactor de cinco factores específicos y un factor general. En todos los casos el valor χ^2 resultó estadísticamente significativo, lo que indica la existencia de una discrepancia entre los modelos propuestos y los datos analizados; sin embargo, dado el tamaño de la muestra, este resultado era esperable (Byrne, 2006).

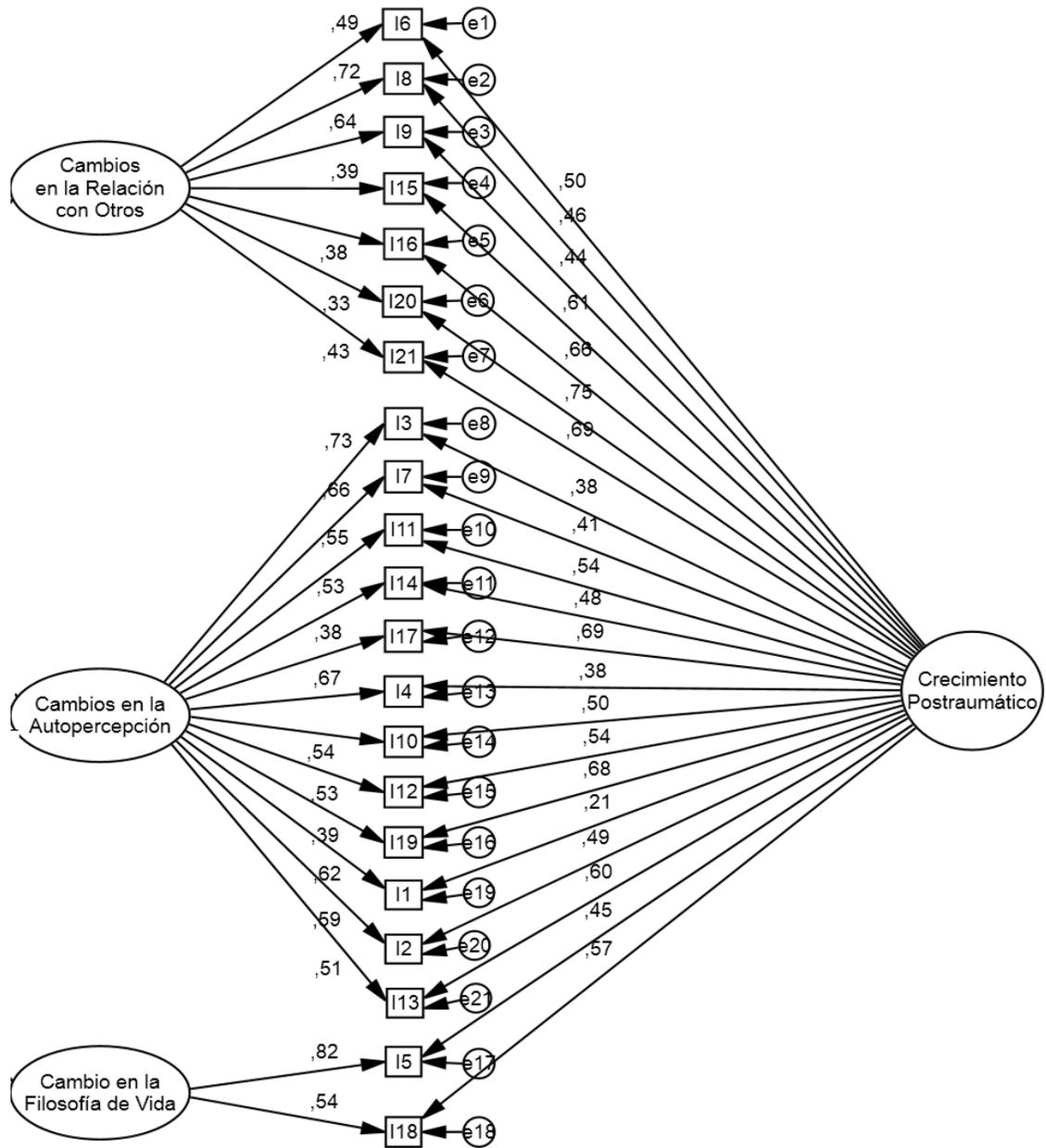
Al analizar los estadísticos CFI, TLI y RMSEA se observa que el modelo de un factor mostró un ajuste insuficiente. Por su parte, los modelos bifactor de tres y cinco factores específicos obtuvieron valores aceptables casi equivalentes, incluso el valor del índice comparativo AIC.

Al analizar las correlaciones entre los factores se observaron diferencias entre el modelo de tres y cinco factores. En el modelo de cinco factores, las correlaciones entre los factores de *apreciación de la vida*, *fuerza personal* y *prioridades de la vida* estuvieron en el rango .96 a .98. En el modelo de tres factores, que supone los factores anteriores agrupados en una sola dimensión de *cambios en la autopercepción*, la más alta

correlación observada fue de .84 entre ese factor y *cambios en la relación con otros*. Estas correlaciones encontradas en el modelo de cinco factores cuestionan su validez discriminante, ya que no son aceptables las correlaciones tan altas entre factores (Brown, 2006). A partir de este criterio, se estableció el modelo bifactor de tres factores específicos y un factor general como el más adecuado. En la Figura 1 se representa este modelo.

Al comparar hombres y mujeres, se observaron diferencias significativas en el PTGI total ($t_{(448)} = 2.30, p = .02$) y obtuvieron mayores puntajes las mujeres que los hombres. A su vez, se observaron diferencias significativas de género tanto en el factor de *cambio en la relación con otros* ($t_{(448)} = 2.89, p = .004$) como en el factor de *cambio en la filosofía de vida* ($t_{(448)} = 3.09, p = .002$), con mayores puntajes en el PTGI por parte de las mujeres. Las medias y desviaciones estándar de cada grupo se muestran en la Tabla 3.

Se procedió a evaluar la invarianza factorial del modelo entre hombres y mujeres. El análisis entre los grupos muestra la existencia de invarianza configural, pues los valores de los índices de ajuste RMSEA, TLI y CFI fueron aceptables. También muestra invarianza métrica, dado que el CFI no disminuyó su valor con respecto al modelo anterior y los otros índices casi no sufrieron

**Figura 1**

Modelo final de tres factores de primer orden y un factor general.

Nota. Los valores están estandarizados.

variaciones, por lo que se puede concluir que los dos grupos presentan cargas factoriales equivalentes. La invarianza fuerte también es aceptable pues los índices de ajuste siguen siendo apropiados y el CFI no disminuye, por lo que es posible concluir que los dos modelos evaluados son equivalentes respecto a los coeficientes factoria-

les y a los interceptos. Finalmente, la invarianza estricta tampoco muestra una disminución en el CFI y presenta valores de ajuste aceptables en los otros índices, mostrando equivalencia en cargas factoriales, interceptos y errores. Con esto se ha aceptado el máximo nivel de invarianza al que se ha sometido el modelo (ver Tabla 4).

Tabla 4

Índices de ajuste para la prueba de invarianza factorial entre hombres y mujeres.

Modelos	$\chi^2 (gl)$	Δ^2	CFI	ΔCFI	TLI	RMSEA
M1: configural	817.59*** (370)	-	.93	-	.92	.05
M2: métrica	827.55*** (388)	-9.96	.93	.00	.93	.05
M3: fuerte	876.32*** (409)	-48.77	.93	.00	.93	.05
M4: estricta	907.88*** (431)	-31.56	.93	.00	.93	.05

Nota. M1 = sin restricciones; M2 = M1 + igualdad en cargas factoriales; M3 = M2 + igualdad en los interceptos; M4 = M3 + igualdad en los errores o residuos. *** $p < .001$.

Los valores del coeficiente alfa de Cronbach de la escala total y de sus factores son adecuados en el rango de .83 a .96 (ver Tabla 3). El rango de la correlación ítem-total corregida fue de .55 a .77 en la escala total, .57 a .77 en *cambios en la autopercepción*, .65 a .78 en *cambios en la relación con otros* y .70 en ambos ítems en *cambios en la filosofía de vida*.

En la Tabla 3 se presentan los estadísticos descriptivos del PTGI tanto para la escala total como para los tres factores específicos obtenidos en el AFC. También se pueden observar las correlaciones entre los distintos factores y la escala total. Al prorratear las medias de cada factor específico y dividir por el número de ítems de dicha subescala, se puede constatar que la dimensión de mayor crecimiento en esta población es la de cambio en la relación con otros (3.14), seguida por la de cambio en la autopercepción (3.07) y finalmente la de cambio en la filosofía de la vida (2.80).

Discusión

En la mayoría de los estudios en los que se ha utilizado AFC para evaluar la estructura factorial del PTGI, se ha observado un modelo de cinco factores similar al del estudio original de

Tedeschi y Calhoun (1996); sin embargo, también se han encontrado modelos de uno y de tres factores.

El presente estudio muestra que el modelo de un factor obtiene un ajuste insuficiente, mientras que los modelos jerárquicos de tres y cinco factores de primer orden y un factor general obtienen valores aceptables casi equivalentes; sin embargo el modelo bifactor de tres factores específicos y un factor general fue el que presentó mejor ajuste a los datos y mayor validez discriminante entre sus distintos factores, por lo que se consideró como el más adecuado. La solución bifactor con un factor general de CPT es concordante con la idea de un modelo teórico subyacente común de CPT (Tedeschi & Calhoun, 1996). Este resultado coincide con lo encontrado por Rodríguez-Rey et al. (2016) y Garrido-Hernansaiz et al. (2017) en población española utilizando una versión más breve del instrumento.

En este modelo, el primer factor específico corresponde a *cambios en la autopercepción*, el segundo a *cambios en la relación con otros* y el tercero a *cambios en la filosofía de vida*, lo que coincide con la propuesta teórica original de Tedeschi y Calhoun (1996), quienes plantearon tres componentes del CPT. Además el presente estudio considera la presencia de un factor general, dando respaldo para considerar al CPT como

un constructo que se puede evaluar tanto a través de sus dimensiones como a través de un puntaje único.

En el análisis de su confiabilidad, la escala mostró una alta consistencia interna tanto a nivel global como en cada uno de sus tres factores. Las correlaciones ítem-total corregidas tanto para la escala total como para los tres factores resultaron adecuadas. Ningún ítem mostró un comportamiento inadecuado.

En este estudio se encontraron diferencias en el CPT entre hombres y mujeres, lo que coincide con lo observado en diversos estudios y en el metaanálisis de Vishnevsky et al. (2010). Se ha planteado como una posible explicación a esta diferencia que las mujeres mantienen un nivel mayor de rumiación de pensamiento en comparación con los hombres tras vivir un hecho estresante, lo que les lleva a experimentar un mayor nivel de sufrimiento y evidencia una mayor prevalencia de estrés postraumático y depresión; pero a su vez, es este mismo sufrimiento el que les permite aprender a medida que se enfrentan a la experiencia (Avilés-Maldonado, Cova-Solar, Bustos, & García, 2014; Vishnevsky et al., 2010).

Se realizó un análisis de la invarianza factorial del modelo comparando hombres y mujeres. Este análisis mostró la existencia de los tres niveles de invarianza evaluados, lo que muestra que los dos grupos presentan cargas factoriales, interceptos y errores equivalentes. Estos resultados permitieron aceptar el máximo nivel de invarianza al que fue sometido el modelo, posibilitando asumir que no existe sesgo por sexo en la escala y que efectivamente existe una diferencia en los niveles de CPT entre ambos grupos.

Dentro del análisis descriptivo se encontró que los sobrevivientes al terremoto del 16 de abril de 2016 mostraron un menor crecimiento en el factor de *cambio en la filosofía de vida* en relación con los otros factores; este fenómeno se

podría explicar considerando que la muestra pertenece a la zona costera de Ecuador, un sector de bajo desarrollo religioso y espiritual en comparación con la zona andina. De la misma forma, los sobrevivientes ecuatorianos presentaron un mayor crecimiento en la dimensión de *cambios en la relación con otros*, fenómeno que se podría explicar por el tipo de sociedad colectivista de los ecuatorianos.

Comparado con una muestra de sobrevivientes chilenos al terremoto del año 2010 del mismo rango de edad (García, Jaramillo, et al., 2014), los puntajes de los participantes ecuatorianos fueron más bajos tanto en sus factores de *cambios en la relación con otros* y *cambios en la autopercepción* como en el puntaje total; sin embargo, el estudio chileno se produjo a los dos años del sismo; en cambio, el presente estudio se realizó doce meses después, por lo que es probable que dicho crecimiento siga aumentando en la medida que las personas tomen distancia suficiente del evento como para considerar áreas de crecimiento a partir de este. Estos resultados deberían ser contrastados con nuevas investigaciones en el campo del CPT para explicar este tipo de comportamiento.

Las limitaciones de este estudio radican en primer lugar en su diseño transversal que impide evaluar la estabilidad temporal de la escala. En segundo término, el muestreo intencionado utilizado puede sugerir que el grupo participante no necesariamente es representativo de la población general del Ecuador. Por último, la no inclusión de otros instrumentos impidió dar un valor discriminante e incluso predictivo a la variable CPT. Surge en consecuencia la posibilidad, al contar ahora con un instrumento psicométricamente adecuado para evaluar el CPT, de realizar estudios que relacionen el crecimiento con otras variables que se han observado relevantes como predictores de la misma. De este modo, sería posible llevar a

cabo otros estudios con población latinoamericana expuesta a un desastre natural, y contrastar los datos obtenidos sobre CPT con variables como la religiosidad (García, Páez-Rovira, Zurtia, Martel, & Reyes, 2014) y el uso de estrategias de afrontamiento (García, Cova, Rincón, Vázquez, & Páez, 2016).

Contar con un instrumento que evalúe la posibilidad de cambios positivos tras vivir un evento adverso abre oportunidades para continuar investigando las secuelas de estos eventos más allá de las respuestas psicopatológicas en las que se ha centrado la investigación en las últimas décadas.

El presente estudio mostró que este instrumento cuenta con propiedades psicométricas apropiadas, lo que implica una alta confiabilidad, una estructura factorial de tres factores específicos y un factor general e invarianza en la medida entre hombres y mujeres, lo que asegura un instrumento no sesgado de acuerdo al sexo. Ante esto, podemos concluir que el PTGI es un instrumento útil para investigar el crecimiento postraumático en población ecuatoriana.

A la vez, esta adaptación permitirá guiar de mejor manera la intervención clínica post desastre al proporcionar una meta objetiva de trabajo terapéutico.

El análisis del CPT en la población afectada tendrá un gran impacto en la salud pública del país, pues permitirá predecir las probabilidades de presencia de respuestas psicopatológicas a largo plazo y desarrollar planes de promoción de la salud mental y prevención.

Por último, la adaptación de este instrumento psicométrico permite sentar una base empírica para la mejora en la gestión de los riesgos naturales centrada en la promoción de una salud mental positiva y en la posibilidad de que una catástrofe pueda ser también una oportunidad de crecimiento personal y colectivo. Esta intervención puede encaminarse hacia la promoción de fortalezas del

carácter y de estrategias de afrontamiento activas alentando la cercanía con la familia y otras redes sociales junto con la generación de mayor valoración de la propia existencia.

Referencias

- American Psychiatric Association (2013). *Manual Diagnóstico y Estadístico los Trastornos Mentales DSM-V*. Barcelona, España: Masson.
- Anderson Jr, W. P., & Lopez-Baez, S. I. (2008). Measuring growth with the Posttraumatic Growth Inventory. *Measurement and Evaluation in Counseling and Development, 40*(4), 215-227.
- Arbuckle, J. L. (2009). *Amos* (Version 20.0) [Computer Program]. Chicago, IL: IBM SPSS.
- Avilés-Maldonado, P., Cova-Solar, F., Bustos, C., & García, F. E. (2014). Afrontamiento y rumiación frente a eventos adversos y crecimiento postraumático en estudiantes universitarios. *Liberabit, 20*(2), 281-292. Recuperado de <http://www.scielo.org.pe/pdf/liber/v20n2/a09v20n2.pdf>
- Bentler, P. (2005). *EQS Structural equations program manual*. Encino, CA: Multivariate Software.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: John Wiley & Sons. doi: 10.1002/9781118619179
- Brown, T. A. (2006). *Confirmatory Factor Analysis for Applied Research*. New York, NY: Guilford Press.
- Brunet, J., McDonough, M. H., Hadd, V., Crocker, P. R. E., & Sabiston, C. M. (2010). The Posttraumatic Growth Inventory: An examination of the factor structure and invariance among breast cancer survivors. *Psycho-Oncology, 19*(8), 830-838. doi: 10.1002/pon.1640
- Byrne, B. (2006). *Structural equation modeling with Mplus*. New York, NY: Routledge.
- Calhoun, L. G., Cann, A., & Tedeschi, R. G. (2010). The Posttraumatic Growth Model: Sociocultural considerations. En T. Weiss & R. Berger (Eds.), *Posttraumatic growth and culturally competent*

- practice: Lessons learned from around the globe* (pp.1-14). New York, NY: John Wiley & Sons. doi: 10.1002/9781118270028.ch1
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9(2), 233-255. doi: 10.1207/S15328007SEM0902_5
- Costa-Requena, G., & Moncayo, F. L. G. (2007). Crecimiento postraumático en pacientes oncológicos. *Análisis y Modificación de Conducta*, 33(148), 229-250. Recuperado de <http://uhu.es/publicaciones/ojs/index.php/amc/article/view/1215>
- Elosúa, P. (2005). Evaluación progresiva de la invarianza factorial entre las versiones original y adaptada de una escala de autoconcepto. *Psicothema*, 17(2), 356-362. Recuperado de: <http://www.ehu.es/gip/publicaciones/articulos/2005/2.pdf>
- Esparza-Baigorria, T., Leibovich de Figueroa, N., & Martínez-Terrer, M. (2016). Propiedades psicométricas del Inventario de Crecimiento Postraumático en pacientes oncológicos en población argentina. *Ansiedad y Estrés* 22(2) 97-103. doi: 10.1016/j.anyes.2016.06.002
- García, F. E., Capponi, C., Hinrichs, F., Lillo, F., Rodríguez, C., & Sánchez, J. (2016). Violencia policial y afrontamiento: Crecimiento postraumático en pobladores de una comunidad aislada del sur de Chile. *Universitas Psychologica*, 15(4). doi: 10.11144/Javeriana.upsy15-4.vpac
- García, F. E., Cova-Solar, F., & Melipillán, R. (2013). Propiedades psicométricas del Inventario de Crecimiento Postraumático en población Chilena afectada por un desastre natural. *Revista Mexicana de Psicología*, 30(2), 143-151. Recuperado de <http://www.redalyc.org/articulo.oa?id=243033029007>
<http://www.redalyc.org/articulo.oa?id=243033029007>
- García, F. E., Cova, F., Rincón, P., Vázquez, C., & Páez, D. (2016). Coping, rumination and posttraumatic growth in people affected by an earthquake. *Psicothema*, 28(1), 59-65. doi: 10.7334/psicothema2015.100
- García, F. E., Jaramillo, C., Martínez, A. M., Valenzuela, I., & Cova-Solar, F. (2014). Respuestas psicológicas ante un desastre natural: Estrés y crecimiento postraumático. *Liberabit*, 20, 121-130. Recuperado de <http://www.scielo.org.pe/pdf/liber/v20n1/a11v20n1.pdf>
- García, F. E., Páez, D., Reyes-Reyes, A., & Álvarez, R. (2017). Religious coping as moderator of psychological responses to stressful events: A longitudinal study. *Religions*, 8(62), 1-13. doi: 10.3390/rel8040062
- García, F. E., Páez-Rovira, D., Zurtia, G. C., Martel, H. N., & Reyes, A. R. (2014). Religious coping, social support and subjective severity as predictors of posttraumatic growth in people affected by the earthquake in Chile on 27/2/2010. *Religions*, 5(4), 1132-1145. doi: 10.3390/rel5041132
- Garcia da Silva, T., Candia-Donat, J., Gauer, G., & Haag-Kristensen, C. (2016). Posttraumatic growth measures: Translation and adaptation of three self-report instruments to Brazilian Portuguese. *Archives of Clinical Psychiatry (Sao Paulo)*, 43(3), 47-50. doi: 10.1590/0101-60830000000083
- Garrido-Hernansaiz, H., Rodríguez-Rey, R., & Alonso-Tapia, J. (2017). Posttraumatic Growth Inventory: Factor structure in Spanish-speaking people living with HIV. *AIDS Care*, 1-4. doi: 10.1080/09540121.2017.1291900
- Hooper, L. M., Marotta, S. A., & Depuy, V. (2009). A confirmatory factor analytic study of the Posttraumatic Growth Inventory among a sample of racially diverse college students. *Journal of Mental Health*, 18(4), 335-343. doi: 10.1080/09638230802522502
- IBM Corp (2011). *IBM SPSS Statistics for Windows*, Version 21.0. Armonk, NY: IBM Corp.
- Joseph, S., & Linley, P. A. (2008). Psychological assessment of growth following adversity: A review. En S. Joseph & P. Linley (Eds.), *Trauma, recovery, and growth: Positive psychological perspectives on posttraumatic stress* (pp. 21-38). New York, NY: John Wiley & Sons. doi: 10.1002/9781118269718.ch2
- Joseph, S., Linley, P. A., & Harris, G. J. (2004). Understanding

- positive change following trauma and adversity: A structural clarification. *Journal of Loss and Trauma*, 10(1), 83-96. doi: 10.1080/15325020490890741
- Kline, R. B. (2005). *Principles and practice of structural equation modeling* (3th ed.). New York, NY: Guilford.
- Konkolý-Thege, B., Kovács, É., & Balog, P. (2014). A bifactor model of the Posttraumatic Growth Inventory. *Health Psychology and Behavioral Medicine: An Open Access Journal*, 2(1), 529-540. doi: 10.1080/21642850.2014.905208
- Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, 57(3), 519-530. doi: 10.2307/2334770
- Mardia, K. V. (1974). Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies. *Sankhya: The Indian Journal of Statistics*, 36(2), 115-128. Recuperado de <http://www.jstor.org/stable/25051892>
- Martínez-Arias, R. (1995). *Psicometría: teoría de los test psicológicos y educativos*. Madrid, España: Síntesis.
- Palmer, G. A., Graca, J. J., & Occhietti, K. E. (2012). Confirmatory factor analysis of the Posttraumatic Growth Inventory in a veteran sample with posttraumatic stress disorder. *Journal of Loss and Trauma*, 17(6), 545-556. doi: 10.1080/15325024.2012.678779
- Powell, S., Rosner, R., Butollo, W., Tedeschi, R. G., & Calhoun, L. G. (2003). Posttraumatic growth after war: A study with former refugees and displaced people in Sarajevo. *Journal of Clinical Psychology*, 59(1), 71-83. doi: 10.1002/jclp.10117
- Rodríguez-Rey, R., Alonso-Tapia, J., Kassam-Adams, N., & Garrido-Hernansaiz, H. (2016). The factor structure of the Posttraumatic Growth Inventory in parents of critically ill children. *Psicothema*, 28(4), 495-503. doi: 10.7334/psicothema2016.162
- Rodríguez-Ayán, M. N., & Ruiz Díaz, M. Á. (2008). Atenuación de la asimetría y de la curtosis de las puntuaciones observadas mediante transformaciones de variables: Incidencia sobre la estructura factorial. *Psicológica: Revista de Metodología y Psicología Experimental*, 29(2), 205-228. Recuperado de <http://www.uv.es/revispsi/articulos2.08/6RODRIGUEZ.pdf>
- Taku, K., Cann, A., Calhoun, L. G., & Tedeschi, R. G. (2008). The factor structure of the Posttraumatic Growth Inventory: A comparison of five models using confirmatory factor analysis. *Journal of Traumatic Stress*, 21(2), 158-164. doi: 10.1002/jts.20305
- Tedeschi, R. G., & Calhoun, L. G. (1996). The Posttraumatic Growth Inventory: Measuring the positive legacy of trauma. *Journal of Traumatic Stress*, 9(3), 455-471. doi: 10.1002/jts.2490090305
- Tedeschi, R. G., & Calhoun, L. G. (2000). Posttraumatic growth: A new focus in psycho traumatology. *Psy-talk, Newsletter of the British Psychological Society Student Members Group*, April 2000.
- Tedeschi, R. G., & Calhoun, L. G. (2008). Crecimiento postraumático en las intervenciones clínicas cognitivo-conductuales. En V. Caballo (Ed). *Manual para el tratamiento cognitivo-conductual de los trastornos psicológicos* (pp. 30-49). Madrid, España: Siglo XXI.
- Vishnevsky, T., Cann, A., Calhoun, L. G., Tedeschi, R. G., & Demakis, G. J. (2010). Gender differences in self-reported posttraumatic growth: A meta-analysis. *Psychology of Women Quarterly*, 34(1), 110-120. doi: 10.1111/j.1471-6402.2009.01546.x
- Weiss, T., & Berger, R. (2006). Reliability and validity of a Spanish version of the Posttraumatic Growth Inventory. *Research on Social Work Practice*, 16(2), 191-199. doi: 10.1177/1049731505281374

Anexo

Inventario de Crecimiento Postraumático

Sexo: F () M () Edad: _____ Fecha: _____

De acuerdo a la escala mostrada a continuación, indique cuánto cambio tuvo en su vida a raíz de la crisis/desastre vivida. Use la siguiente escala:

0: No he experimentado ningún cambio en mi vida a raíz de la crisis/desastre

1: Ha habido un cambio minúsculo en mi vida a raíz de la crisis/desastre

2: Ha habido un pequeño cambio en mi vida a raíz de la crisis/desastre

3: Ha habido un cambio moderado en mi vida a raíz de la crisis/desastre

4: Ha habido un gran cambio en mi vida a raíz de la crisis/desastre

5: Ha habido un cambio enorme en mi vida a raíz de la crisis/desastre

Posibles áreas de crecimiento y cambio		0	1	2	3	4	5
1	Han cambiado mis prioridades acerca de lo que es importante en mi vida.						
2	Tengo una mejor apreciación acerca del valor de mi propia vida.						
3	Desarrollé nuevos intereses.						
4	Tengo un mejor sentido de autosuficiencia.						
5	Soy una persona más espiritual, he crecido espiritualmente.						
6	Puedo apreciar más claramente que puedo contar con los demás en tiempos difíciles.						
7	He construido un nuevo rumbo o caminos de vida.						
8	Siento un mayor sentido de proximidad con las personas o mayor cercanía hacia los demás.						
9	Estoy más dispuesto a expresar mis emociones.						
10	Siento que puedo arreglármelas mejor en los momentos difíciles.						
11	Puedo hacer mejores cosas con mi vida.						
12	Estoy más dispuesto a aceptar la forma en que las cosas se dan.						
13	Puedo apreciar mejor cada día.						
14	Hay nuevas oportunidades que antes no estaban disponibles.						
15	Siento más compasión por los demás.						
16	Pongo más esfuerzo en mis relaciones.						
17	Estoy más dispuesto a intentar cambiar las cosas que necesitan ser cambiadas.						
18	Tengo o he desarrollado una mayor fortaleza religiosa.						
19	He descubierto que era o soy más fuerte de lo que pensaba.						
20	He aprendido lo maravillosas que son las personas o lo extraordinarias que pueden ser.						
21	Acepto de mejor manera el necesitar a los demás.						

Análisis Factorial Confirmatorio del IPAM en Escolares de Tercer Curso de Primaria

Confirmatory Factor Analysis of IPAM in Third-Grade Students

Juan E. Jiménez *¹, Sara del Cristo de León¹

1 - Universidad de La Laguna, España.

Recibido: 12/07/2017 **Revisado:** 16/08/2017 **Aceptado:** 16/08/2017

Introducción
Método
Resultados
Discusión
Referencias

Resumen

Este estudio pretende evaluar la estructura factorial del instrumento Indicadores de Progreso de Aprendizaje en Matemáticas (IPAM) mediante la técnica de análisis factorial confirmatorio (AFC). Con este fin, se ha llevado a cabo un estudio longitudinal con una muestra de 234 alumnos de tercer curso de educación primaria de las Islas Canarias, a los que se administró el instrumento IPAM, un instrumento de medición basado en el currículo (CBM, por sus siglas en inglés, *curriculum-based measurement*), y cuyo principal objetivo es el cribado universal y la evaluación del progreso en el aprendizaje en matemáticas del alumnado de educación primaria. Este instrumento está compuesto por tres medidas paralelas (A, B y C), que pretenden medir una misma estructura latente, el *sentido numérico*, por medio de la resolución de cinco tareas de fluidez (comparación numérica, operaciones de dos dígitos, series numéricas, operaciones de un dígito y valor de posición). El IPAM fue aplicado en tres momentos diferentes a lo largo del año escolar (i.e., otoño, invierno y primavera) y los resultados del AFC mostraron un buen ajuste del modelo propuesto en los distintos momentos de medida.

Palabras clave: *sentido numérico, educación primaria, matemáticas, análisis factorial confirmatorio, medida basada en el currículo*

Abstract

This study has been designed to analyse the factorial structure of IPAM using Confirmatory Factorial Analysis (CFA) techniques. For this purpose, a longitudinal study was carried out with a sample of 234 third-grade elementary students from the Canary Islands, to whom the instrument IPAM (Mathematics Learning Progress Indicators) was administered. IPAM is a curriculum-based measurement (CBM) instrument for universal screening and mathematics learning progress monitoring in elementary grades. It is composed by three parallel measurements (A, B and C) that aim to measure the same latent structure (i.e., *number sense*) through the assessment of five indicators of basic early math skills using fluency tasks (i.e., magnitude comparison, two-digit operations, missing number, one-digit operations, position value). IPAM was administered three times throughout the school year (i.e., fall, winter, and spring). The model tested showed a good fit at the different moments of measurement.

Keywords: *number sense, elementary education, mathematics, confirmatory factor analysis, curriculum-based measurement*

***Correspondencia a:** Juan E. Jiménez. Dirección Postal: Facultad de Psicología, Campus de Guajara, S/N 38200, La Laguna, Islas Canarias. Tel.: +34 922 317545. ejimenez@ull.es

Cómo citar este artículo: Jiménez, J. E., & De León, S. del C. (2017). Análisis factorial confirmatorio del IPAM en escolares de tercer curso de primaria. *Revista Evaluar*, 17(2), 81-96. Recuperado de <https://revistas.unc.edu.ar/index.php/revaluar>

Nota de autor: Este trabajo ha sido posible gracias a la financiación del Plan Nacional I+D+i del Ministerio de Economía y Competitividad, con ref. EDU2012-35098. Investigador Principal: Juan E. Jiménez.

Introducción

El sentido numérico ha demostrado ser una habilidad fundamental en el aprendizaje de las matemáticas (Dyson, Jordan, & Glutting, 2011; Geary, 2013; Jordan, Glutting, & Ramineni, 2010; Jordan, Kaplan, Ramineni, & Locuniak, 2009; Kolkman, Kroesbergen, & Leseman, 2013; Libertus, Fiegenson, & Halberda, 2011; Lyons & Beilock, 2011; National Council of Teachers of Mathematics [NCTM], 2000; Park & Brannon, 2014; Sasanguie, Göbel, Moll, Smets, & Reynvoet, 2013; Sayers & Andrews, 2015; Starr, Libertus, & Brannon, 2013; Toll, Kroesbergen, & Van Luit, 2016). Este constructo ha sido analizado desde una perspectiva evolutiva y se ha llegado a diferenciar un sentido numérico de naturaleza más informal, previo a la propia adquisición del habla (Dehaene, 2009; Jordan & Levine, 2009; Libertus et al., 2011; Piazza, 2010), y un sentido numérico de carácter formal, que se va desarrollando por medio de la escolarización (Andrews & Sayers, 2015; Jordan & Levine, 2009; National Mathematic Advisory Panel [NMAP], 2008).

Dehaene (2009) propuso que la primera fase del desarrollo del sentido numérico viene “incorporada” en el ser humano, lo que nos permite determinar aproximadamente el número de objetos que hay en una escena, si ese número es mayor o menor que otro, y cómo cambia, por medio de adiciones y sustracciones sencillas, de manera rápida, automática e inconsciente. El sentido numérico evoluciona de manera progresiva y posibilita la adquisición de un sentido numérico simbólico, que nos permite comprender y manipular los números y su representación escrita (Dehaene, 2009). El inicio de la escolarización facilita la formación del *sentido numérico fundamental o básico* (FoNS, por sus siglas en inglés, *Foundational Number Sense*), el cual se construye sobre las habilidades subyacentes al *sentido numérico prever-*

bal (Andrews & Sayers, 2015; Jordan & Levine, 2009; Sayers & Andrews, 2015).

Son muchos los autores que han aportado evidencia de los componentes que subyacen al FoNS. En este sentido, Andrews y Sayers (2015) han establecido al menos ocho habilidades básicas: reconocimiento del número, sistema de conteo, relación entre los números y las cantidades que estos representan, discriminación de cantidades, reconocimiento de las distintas representaciones del número, estimación, operaciones aritméticas básicas y patrones numéricos. En consonancia con esta clasificación, otros autores han encontrado que las habilidades vinculadas al conteo, las relaciones entre los números y la resolución de operaciones aritméticas son clave en el desarrollo del sentido numérico (Dyson et al., 2011; Jordan et al., 2010; Jordan & Levine, 2009). En la misma línea, el NMAP (2008) ha establecido que el sentido numérico supone disponer de habilidades de conteo, discriminación de cantidades, estimación de magnitudes, resolución de operaciones aritméticas sencillas, y comprensión del valor de posición de los dígitos dentro de los números.

Atendiendo a lo expuesto, en este trabajo se han considerado cinco categorías para definir las habilidades subyacentes al sentido numérico: 1) magnitudes numéricas; 2) conteo; 3) línea numérica; 4) valor de posición, y 5) operaciones aritméticas.

¿Cómo medir el sentido numérico en tercer curso de educación primaria?

Las medidas basadas en el currículo (CBM, por sus siglas en inglés, *curriculum based measurement*) han demostrado ser adecuadas para establecer diferencias en el desempeño académico entre los estudiantes (Foegen, Jiban, & Deno, 2007; Gersten et al., 2012; Jitendra, Dupuis, &

Zaslowsky, 2014), lo que facilita la detección temprana de estudiantes en riesgo de presentar dificultades de aprendizaje en el área de las matemáticas (Lembke & Foegen, 2009; Peng et al., 2016; Sasanguie et al. 2013; Sisco-Taylor, Fung, & Swanson, 2015). Un gran número de herramientas basadas en el sentido numérico han demostrado poseer un potencial adecuado para la evaluación del logro académico en esta área (Andrews & Sayers, 2015; Jordan et al., 2010; Lembke & Foegen, 2009; Schneider et al., 2016). No obstante, muchas de las investigaciones llevadas a cabo, centradas en las habilidades básicas que subyacen al sentido numérico (Kim, Shin, & Lee, 2013; Kolkman et al., 2013; Jiménez & De León, 2017; Lyons & Beilock, 2011; Sasanguie et al. 2013) se han enfocado en los primeros cursos de la escolaridad (i.e., educación infantil y primer y segundo curso de educación primaria; Andrews & Sayers, 2015; Dyson et al., 2011; Foegen et al., 2007; Gersten et al., 2012; Hassinger-Das, Jordan, Glutting, Irwin, & Dyson, 2014; Jiménez & De León 2017; Jordan et al., 2010; Lembke & Foegen, 2009; Libertus, Fiegenson, & Halberda, 2013; Sasanguie et al., 2013; Starr et al., 2013; Toll et al., 2016), mientras que es menor el número de investigaciones llevadas a cabo en cursos superiores (i.e., tercero, cuarto, quinto y sexto año de educación primaria; Foegen et al., 2007; Jitendra et al., 2014; Jordan et al., 2010; Lyons, Price, Vaessen, Blomert, & Ansari, 2014; Namkung & Fuchs, 2016).

Considerando lo expuesto anteriormente, se ha diseñado la medida Indicadores de Progreso de Aprendizaje en Matemáticas (IPAM; Jiménez & De León, 2016), que incluye un total de cinco tareas: comparación numérica, secuencias numéricas, operaciones de uno y dos dígitos, y valor de posición.

El objetivo de este trabajo es poner a prueba un modelo hipotético que permita analizar la es-

tructura factorial de este instrumento en alumnos de tercer curso de educación primaria. Para ello se empleará la técnica de análisis factorial confirmatorio (AFC), con el fin de conocer en qué medida la varianza asociada a cada una de las tareas que componen la herramienta IPAM (comparación numérica, operaciones de dos dígitos, series numéricas, operaciones de un dígito, y valor de posición) es explicada por un único factor asociado al constructo *sentido numérico*.

Método

Participantes

La población de este estudio corresponde al alumnado de 3º curso de educación primaria perteneciente a cinco centros de educación primaria, uno concertado y cuatro públicos, de la provincia de Santa Cruz de Tenerife. La muestra incluyó un total de 234 estudiantes, de los cuales 100 eran niños y 134 eran niñas. La edad promedio de la muestra fue de 8.34, con un rango de edad de 7.33 a 9.75 años.

Instrumentos

Se utilizó la herramienta Indicadores de Progreso de Aprendizaje en Matemáticas (IPAM; Jiménez & De León, 2016) diseñada para tercer curso de educación primaria. El IPAM es un instrumento CBM compuesto por tres medidas alternativas o paralelas (A, B y C) que han sido diseñadas para medir un mismo constructo (el sentido numérico). Estas medidas se aplican de manera trimestral, en noviembre (Forma A), febrero (Forma B), y mayo (Forma C) respectivamente. El principal objetivo de este instrumento es el cribado universal y la evaluación del progreso en el aprendizaje del alumnado, mediante la

evaluación de su fluidez en la resolución de cada tarea. La fluidez es medida contabilizando el número de respuestas correctas que da el alumno en un tiempo dado.

Cada una de las formas está compuesta por cinco tareas, que se aplican de forma consecutiva con una duración máxima de dos minutos por tarea. En la primera tarea, comparación numérica (CN), al estudiante se le presenta una serie de parejas de números y se le solicita rodear aquel que presente mayor magnitud en cada una de las parejas dadas. En la segunda tarea, operaciones de dos dígitos (ODD), se requiere que el estudiante resuelva de forma correcta operaciones de adición, sustracción y multiplicación. En la tercera tarea, secuencias numéricas (SN), se presenta una serie de secuencias compuestas por dos dígitos y una incógnita, la cual el alumno debe completar con el número correspondiente. La cuarta tarea, operaciones de un dígito (OUD), implica la resolución de operaciones básicas de adición, sustracción y multiplicación, en las que se emplean números del 1 al 9. Finalmente, en la quinta tarea, valor de posición (VP), a partir de una representación gráfica dada el estudiante debe determinar el valor posicional de las cifras.

Procedimiento

Antes de la administración del instrumento, se solicitó autorización a cada centro educativo y a las familias para la realización del estudio. Para llevar a cabo la recogida de información siete licenciados en Psicología y dos licenciados en Psicopedagogía fueron entrenados en la aplicación de la herramienta IPAM.

Las pruebas se aplicaron en tres momentos durante el año escolar (noviembre, febrero y mayo). La aplicación se llevó a cabo de forma colectiva en cada una de las aulas de clase y en horario

lectivo, y se informó a los estudiantes del objetivo del estudio.

Análisis de datos

En este estudio se postula un modelo y se lo evalúa mediante un AFC efectuado sobre los datos empíricos obtenidos a través de las respuestas proporcionadas por los estudiantes en la herramienta IPAM. El modelo propuesto postula la existencia de un único factor de primer orden y, por tanto, que todas las tareas que incluye la herramienta están relacionadas con ese factor. Este modelo se corresponde con una concepción teórica unidimensional del sentido numérico.

En la primera fase hemos formulado la estructura de relaciones entre los indicadores observables y el factor latente. Una vez establecida la estructura del instrumento a nivel conceptual, ésta se postula como la teoría cuya dimensionalidad será validada, con base en los resultados obtenidos.

El segundo paso ha consistido en emplear la metodología de modelado de ecuaciones estructurales (SEM, por sus siglas en inglés, *Structural Equation Modeling*) implementada mediante el paquete LISREL 8.8, con la modalidad de AFC (Jöreskog & Sörbom, 1996-2001). Este procedimiento consiste en analizar la estructura de covarianza en la base de datos que contiene los cinco indicadores observables (tareas) para obtener evidencia de validez de constructo y poder evaluar si los datos se ajustan al modelo propuesto.

La fase de especificación del modelo incluye una representación gráfica de la estructura teórico-conceptual del instrumento bajo análisis. Por ser el IPAM una herramienta formada por tres medidas, se presenta una representación gráfica de cada una de ellas en el análisis de los resultados en las Figuras 1-3.

En cada una de estas figuras, el círculo representa el factor latente (constructo) y los óvalos, los cinco indicadores observables (tareas). Las flechas unidireccionales representan las líneas de influencia del factor sobre los respectivos indicadores observables. Estas especificaciones de direccionalidad equivalen a afirmar que cada uno de los indicadores relacionados mide la dimensión a la que en teoría pertenece. Finalmente, las flechas unidireccionales ubicadas a la derecha de los indicadores observables representan los términos de error, que son la influencia de fuentes de variabilidad única (o idiosincrática) asociadas a cada indicador observable.

Se utilizó el Software R (R Core Team, 2013) para el cálculo de estadísticos descriptivos y de los coeficientes de correlación de los resultados obtenidos en las tareas, los cuales se calcularon con el fin de confirmar la ausencia de cualquier efecto de multicolinealidad entre los indicadores. Se calculó la media, desviación típica, máximo, mínimo, asimetría y curtosis para caracterizar los datos de la muestra.

Mediante el Software R y ULLRToolbox (Hernández & Betancort, 2016), se evaluó la validez factorial de este modelo a través de un AFC. La confiabilidad compuesta (*Composite Reliability*) y la varianza media extraída (*Average Variance Extracted*) para cada factor se calcularon según las indicaciones de Fornell y Larcker (1981). La existencia de los valores atípicos se evaluó por la distancia al cuadrado de Mahalanobis (D^2) y la normalidad de los indicadores se evaluó mediante los coeficientes de asimetría y curtosis, uni y multivariante.

Se utilizó el método de estimación de máxima verosimilitud. La evaluación del método se realizó mediante los siguientes índices: chi-cuadrada (χ^2), chi cuadrada sobre los grados de libertad (χ^2/df), índice de ajuste comparativo (CFI), bondad de ajuste del modelo (GFI), y error

cuadrático medio de aproximación (RMSEA), P [RMSEA \leq .05].

Resultados

A la hora de exponer los resultados es importante recordar que la herramienta IPAM está compuesta por tres formas aplicadas consecutivamente a intervalos regulares. De este modo, se expondrán en primer lugar los resultados correspondientes a la Forma A, en segundo lugar, los resultados de la Forma B y, finalmente, los resultados de la Forma C.

IPAM Forma A

En la Tabla 1 se presentan los análisis estadísticos descriptivos de cada una de las tareas que componen el IPAM (Forma A). Se muestran los valores máximo y mínimo, medias, desviación estándar, asimetría y curtosis para cada tarea. Un primer análisis de estos datos muestra que los valores de asimetría y curtosis son cercanos a cero. De este modo, se evidencia una distribución normal y una dispersión satisfactoria de los resultados, con la excepción de *valor de posición*, que tiende a valores más altos en relación al promedio, aunque se mantiene dentro de los rangos aceptables para estos estadísticos. Al observar los valores de asimetría y curtosis de todas las variables del estudio se puede corroborar que no existen problemas con la distribución de los datos, puesto que los valores absolutos de estos índices no exceden los valores 3.0 y 10.0 respectivamente (Kline, 2005). Las desviaciones en las puntuaciones estándar se sitúan entre 9.98 y 2.76.

Tabla 1
Estadísticos descriptivos de IPAM Forma A.

	Mínimo	Máximo	Media	Desviación estándar	Asimetría	Curtosis
CN	4	63	39.35	9.98	-.40	1.23
ODD	.0	16	5.52	2.76	.46	.82
SN	.0	22	8.59	5.65	.18	-1.02
ODD	.0	40	13.52	6.17	.30	-.21
VP	.0	16	6.89	3.36	1.14	2.98

Nota. CN = Comparación numérica, ODD = Operaciones de dos dígitos, SN = Series numéricas, OUD = Operaciones de un dígito, VP = Valor de posición.

Tabla 2
Coeficientes de correlación entre las tareas de IPAM Forma A.

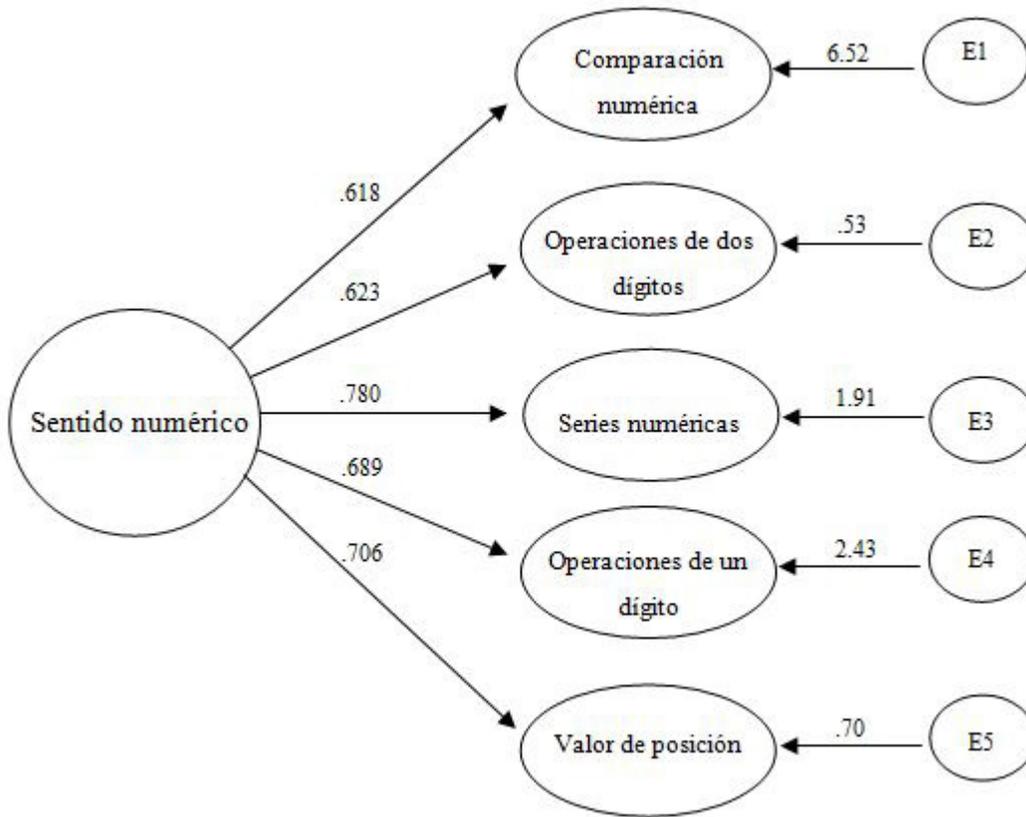
	CN	ODD	SN	ODD	VP
CN	1.00	.33***	.47***	.43***	.48***
ODD		1.00	.51***	.60***	.44***
SN			1.00	.55***	.54***
ODD				1.00	.47***
VP					1.00

Nota. CN = Comparación numérica, ODD = Operaciones de dos dígitos, SN = Series numéricas, OUD = Operaciones de un dígito, VP = Valor de posición.
*** $p < .001$.

La matriz de correlación de los resultados obtenidos en las diferentes tareas que componen el IPAM (Forma A) se presenta en la Tabla 2. Los resultados obtenidos muestran correlaciones estadísticamente significativas entre todos los indicadores.

En primer lugar, el AFC da como resultado una representación de las relaciones estructurales observadas en la matriz derivada. Los valores que representan las flechas unidireccionales son las líneas de influencia entre el factor latente y los indicadores observables y se interpretan como coeficientes de regresión multivariada. Estos han sido calculados por el programa, al analizar la matriz de covarianza derivada de los indicadores observables. Así también, se han calculado los coeficientes de regresión entre el factor latente y

los indicadores observables. El primer valor que se encuentra es .61, el cual corresponde a la influencia del factor latente *sentido numérico* sobre el indicador observable *comparación numérica* (CN). El coeficiente de regresión presenta signo positivo, lo que evidencia una relación directa entre el factor latente y el indicador observable. Esto implica que cuando el factor latente aumenta una unidad, la variable observada CN aumenta en una proporción de .61. La misma interpretación cabe para las relaciones existentes entre el factor latente y el resto de los indicadores observables, ya que se encontró un coeficiente de .63 para *operaciones con dos dígitos* (ODD), un coeficiente de .78 para *series numéricas* (SN), un coeficiente de .68 para *operaciones con un dígito* (OUD), y un coeficiente de .70 para *valor de posición* (VP) (ver Figura 1).

**Figura 1**

Modelo de un factor general. IPAM Forma A

La evaluación preliminar del modelo permitió verificar en los coeficientes derivados el sentido positivo esperado de acuerdo con la teoría empleada en la configuración del modelo de medida.

Posteriormente, se analizó el nivel de ajuste del modelo propuesto con los datos. En primer lugar, se obtuvo un valor de χ^2 igual a 4.28 ($p = .37$), lo que implica que el modelo y los resultados se ajustan entre sí. Para confirmar este resultado, se procedió a la evaluación del ajuste del modelo mediante los índices de bondad de ajuste. Los resultados indicaron un buen ajuste del modelo a los datos empíricos (SRMR = .02; $\chi^2 = 4.28$; $df = 4$; $p = .37$; NFI = .99; CFI = .99; GFI = .99; AGFI = .97; RMSEA = .02; IC = .00 - .10).

Para conseguir un ajuste adecuado del modelo fue necesario liberar la covarianza de error

entre los indicadores observables ODD y OUD. Una posible característica común compartida por estos indicadores y no explicada por la variable latente estaría relacionada con la capacidad para inhibir respuestas competitivas y seleccionar la respuesta más adecuada (LeFevre et al., 2013; Lemaire & Lecacheur, 2011; Lucangeli, Tressoldi, Bendotti, Bonanomi, & Siegel, 2003). Por ejemplo, cuando un niño resuelve una suma, puede utilizar varias estrategias, como contar desde el primer número de la operación en adelante, contar desde el mayor, o recuperar el resultado de un hecho numérico ya almacenado en la memoria (Siegler & Araya, 2005). Por otro lado, cuando el niño ya ha asimilado determinados hechos numéricos, pero aún está aprendiendo otros, los que ya conoce se activan al intentar resolver una

Tabla 3
Estadísticos descriptivos de IPAM Forma B.

	Mínimo	Máximo	Media	Desviación estándar	Asimetría	Curtosis
CN	11	64	47.29	10.53	-.41	.20
ODD	0	21	7.41	3.23	.73	1.66
SN	0	27	10.62	5.94	.13	-.70
ODD	0	45	15.76	7.74	.91	1.49
VP	0	20	7.26	4.05	.45	.13

Nota. CN = Comparación numérica, ODD = Operaciones de dos dígitos, SN = Series numéricas, OUD = Operaciones de un dígito, VP = Valor de posición.

Tabla 4
Coeficientes de correlación entre las tareas de IPAM Forma B.

	CN	ODD	SN	ODD	VP
CN	1.00	.36***	.46***	.43***	.46***
ODD		1.00	.58***	.69***	.50***
SN			1.00	.66***	.66***
ODD				1.00	.63***
VP					1.00

Nota. CN = Comparación numérica, ODD = Operaciones de dos dígitos, SN = Series numéricas, OUD = Operaciones de un dígito, VP = Valor de posición.
.*** $p < .001$.

determinada operación. Por ello, es necesario que inhiba las respuestas que se activan automáticamente por estar asociadas a un determinado hecho numérico conocido, pero que son incorrectas. Se trataría, por ejemplo, de inhibir la respuesta “8” ante la pregunta “¿cuánto es 4×4 ?”, ya que “8” es la respuesta automática que se activa, por tener asimilado que “ $4 + 4$ es igual a 8” (Cragg & Gilmore, 2014; Siegler & Araya, 2005).

Tanto en OUD como en ODD parece evidente que la atención ejecutiva, a la hora de inhibir determinada información en favor de otra más adecuada, así como de seleccionar la estrategia que mejor se ajusta a la demanda de la tarea, puede jugar un papel importante en la resolución de estas tareas, ya que el estudiante debe resolver correctamente sumas, restas y multiplicaciones en el menor tiempo posible, lo que le obliga a se-

leccionar una estrategia que le permita actuar de forma rápida y eficaz, es decir, con fluidez. En este sentido, LeFevre et al. (2013) encontraron que la atención ejecutiva es un factor crítico en la resolución de operaciones aritméticas de manera fluida y en el acceso a los hechos numéricos.

IPAM Forma B

En la Tabla 3 se presentan los análisis estadísticos descriptivos de cada una de las tareas que componen el IPAM (Forma B). Se muestran los valores máximo y mínimo, medias, desviación estándar, asimetría y curtosis para cada tarea. Un primer análisis de estos datos muestra una dispersión satisfactoria para todas las tareas. Los coeficientes de curtosis se aproximan a cero, lo que

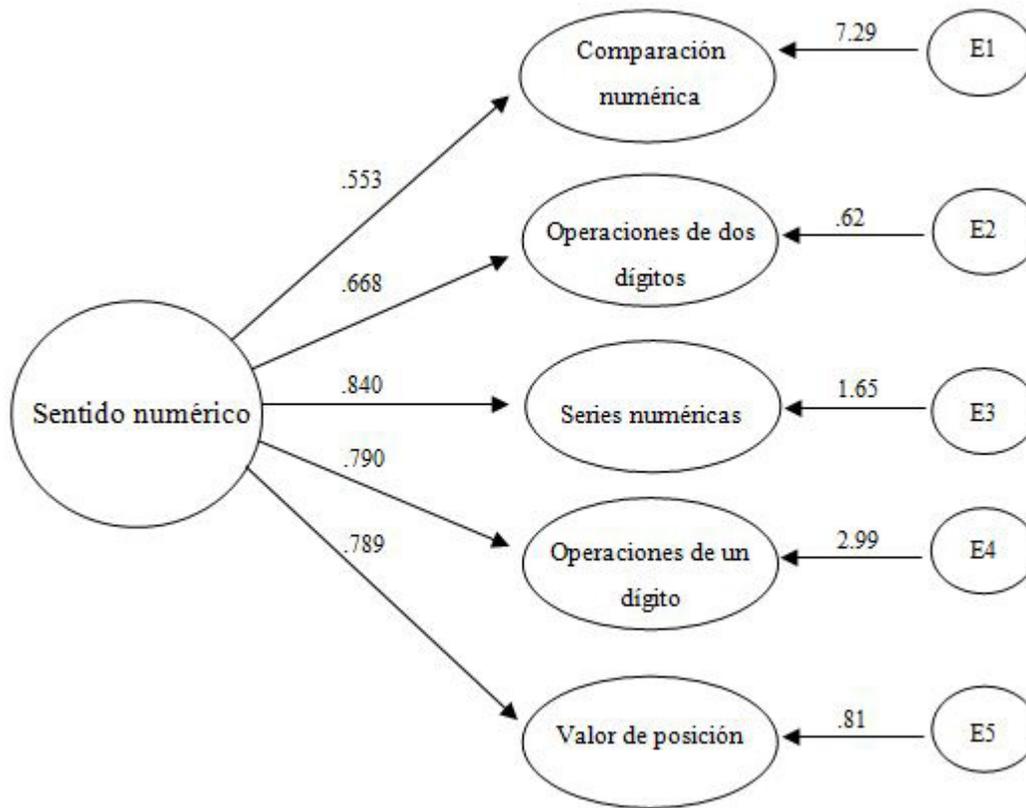


Figura 2
Modelo de un factor general. IPAM Forma B

indica una distribución normal de los resultados (Kline, 2005). Las desviaciones en las puntuaciones estándar se sitúan entre 10.53 y 3.23.

La matriz de correlación de los resultados obtenidos en las diferentes tareas que componen el IPAM (Forma B) se presenta en la Tabla 4. Los resultados obtenidos muestran correlaciones estadísticamente significativas entre todos los indicadores.

Los resultados obtenidos en el AFC de la Forma B del IPAM se han interpretado de la misma manera que aquellos de la Forma A, en cuanto a la representación de las relaciones estructurales observadas en la matriz derivada y a los coeficientes de regresión multivariada. De este modo, el sentido numérico funciona como factor latente y las tareas son sus indicadores observables. En

este caso, CN presenta un coeficiente de regresión de .55, ODD un coeficiente de .66, SN un coeficiente de .84, OUD un coeficiente de .79 y VP presenta un coeficiente de .78 (ver Figura 2).

Así, la evaluación preliminar del modelo permitió verificar para los coeficientes derivados el sentido positivo esperado de acuerdo con la teoría empleada en la configuración del modelo de medida.

Posteriormente, se analizó el nivel de ajuste del modelo propuesto con los datos. El valor de $\chi^2 = 2.06$ ($p = .72$) implica que el modelo y los resultados se ajustan entre sí. Los índices de bondad de ajuste confirman el ajuste del modelo (SRMR = .01; $\chi^2 = 2.06$; $df = 4$; $p = .72$; NFI = .99; CFI = 1; GFI = .99; AGFI = .98; RMSEA = 0; IC = 0 - .07).

Tabla 5
Estadísticos descriptivos de IPAM Forma C.

	Mínimo	Máximo	Media	Desviación estándar	Asimetría	Curtosis
CN	22	64	52.31	9.55	-.78	.22
ODD	0	27	7.27	4.00	1.30	3.58
SN	0	28	12.19	6.26	-.09	-.53
OUD	1	45	21.05	8,57	.44	.32
VP	0	44	9.97	6.13	1.65	5.65

Nota. CN = Comparación numérica, ODD = Operaciones de dos dígitos, SN = Series numéricas, OUD = Operaciones de un dígito, VP = Valor de posición.

Tabla 6
Coeficientes de correlación entre las tareas de IPAM Forma C.

	CN	ODD	SN	OUD	VP
CN	1.00	.47***	.50***	.58***	.41***
ODD		1.00	.60***	.73***	.38***
SN			1.00	.66***	.49***
OUD				1.00	.43***
VP					1.00

Nota. CN = Comparación numérica, ODD = Operaciones de dos dígitos, SN = Series numéricas, OUD = Operaciones de un dígito, VP = Valor de posición.
*** $p < .001$.

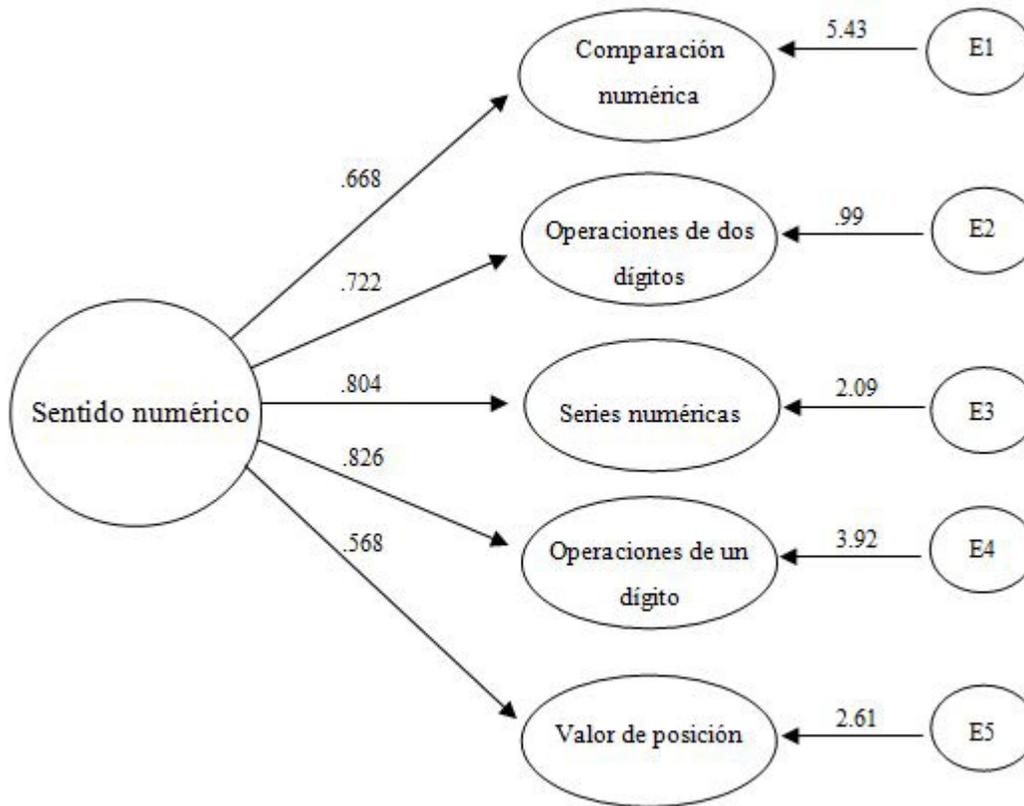
Para que este modelo se ajustase a los datos fue también necesario liberar la covarianza de error existente entre ODD y OUD. Esta covarianza, al igual que en la forma A del IPAM, estaría explicada por la capacidad de inhibir distintas estrategias que se activarían cuando se trata de resolver una determinada operación (LeFevre et al., 2013; Lemaire & Lecacheur, 2011; Lucangeli et al., 2003).

IPAM Forma C

En la Tabla 5 presentamos los análisis estadísticos descriptivos de cada una de las tareas que componen el IPAM (Forma C). Se muestran los valores máximo y mínimo, medias, desviación estándar, asimetría y curtosis para cada tarea. Un primer análisis de estos datos muestra que los

resultados presentan una dispersión satisfactoria para todas las tareas salvo para ODD y VP. Los valores de asimetría y curtosis son cercanos a cero, lo que indica una distribución normal de los resultados, con la excepción de ODD y VP, los cuales presentan valores por encima de la unidad/uno. A pesar de ello, podemos observar que no existen problemas con la distribución de los datos, puesto que los valores absolutos de estos estadísticos no exceden los valores 3.0 y 10.0 respectivamente (Kline, 2005). Las desviaciones en las puntuaciones estándar se sitúan entre 9.55 y 4.00.

La matriz de correlación de los resultados obtenidos en las diferentes tareas que componen el IPAM (Forma C) se presenta en la Tabla 6. Los resultados obtenidos muestran correlaciones estadísticamente significativas entre todos los indicadores.

**Figura 3**

Modelo de un factor general. IPAM Forma C

Los resultados obtenidos en el AFC de la Forma C del IPAM se han interpretado de la misma manera que aquellos de la Forma A, en lo que se refiere a la representación de las relaciones estructurales observadas en la matriz derivada y a los coeficientes de regresión multivariada. De este modo, el sentido numérico funciona como factor latente y las tareas son sus indicadores observables. En este caso, CN presenta un coeficiente de regresión de .66, ODD un coeficiente de .72, SN un coeficiente de .80, OUD un coeficiente de .82 y VP presenta un coeficiente de .56 (Ver Figura 3).

Así, la evaluación preliminar del modelo permite verificar que los coeficientes derivados presentan el sentido positivo esperado de acuerdo con la teoría empleada en la configuración del modelo de medida.

Posteriormente, se analizó el nivel de ajuste del modelo propuesto con los datos. El valor de $\chi^2 = 7.77$ ($p = .101$) implica que el modelo y los resultados se ajustan entre sí. Los índices obtenidos a través del análisis de la bondad de ajuste confirman el ajuste del modelo (SRMR = .02; $\chi^2 = 7.77$; $df = 4$; $p = .10$; NFI = .98; CFI = 1; GFI = .99; AGFI = .95; RMSEA = .06; IC = 0 - .13).

Para que este modelo se ajustase a los datos fue necesario liberar la covarianza de error existente entre ODD y OUD. Esta covarianza, al igual que en las formas A y B del IPAM, estaría explicada por la capacidad para inhibir respuestas competitivas y seleccionar la respuesta más adecuada (LeFevre et al., 2013; Lemaire & Lecacheur, 2011; Lucangeli et al., 2003). Esta habilidad, dependiente de las funciones ejecuti-

vas (es decir, inhibición y flexibilidad cognitiva o atención ejecutiva), permite a los estudiantes resolver operaciones aritméticas con mayor fluidez (LeFevre et al., 2013).

Discusión

En este trabajo se ha pretendido poner a prueba un modelo hipotético para verificar la estructura factorial del instrumento IPAM, diseñado para medir el sentido numérico, en tercer curso de educación primaria, mediante la técnica de análisis factorial confirmatorio (AFC). Los datos obtenidos y el análisis de la estructura factorial según la técnica AFC permiten verificar un buen ajuste del modelo propuesto con los datos empíricos. Así, se observa cómo el factor latente *sentido numérico* explica cada uno de los indicadores observables (comparación numérica, operaciones de dos dígitos, secuencias numéricas, operaciones de un dígito, valor de posición), para cada uno de los momentos de medida (noviembre, febrero y mayo).

Los resultados obtenidos se ajustan a lo prescrito por la teoría en cuanto a la definición del constructo y las habilidades que lo conforman, incluidas las habilidades tempranas típicamente asociadas con el desarrollo del sentido numérico (Andrews & Sayers, 2015; Dyson et al., 2011; Jordan et al., 2010; Jordan & Levine, 2009; NMAP, 2008; Sayers & Andrews, 2015).

Varios trabajos han sido desarrollados con el fin de diseñar herramientas efectivas, basadas en el sentido numérico, que permitan evaluar el rendimiento en matemáticas de los estudiantes de tercer curso de educación primaria. En este sentido, Lyons et al. (2014) analizaron la relación existente entre las habilidades de procesamiento simbólico del número y las habilidades aritméticas en alumnos de primero a sexto curso de edu-

cación primaria. Los resultados mostraron que las habilidades de las que disponían los niños en procesamiento numérico explicaban la mayoría de la varianza asociada a las habilidades aritméticas en estos cursos. Schneider et al. (2016) encontraron que la habilidad para procesar magnitudes (cálculo mental y habilidades matemáticas tempranas) estaba altamente asociada con las competencias matemáticas de las que disponían los estudiantes, y era una de las habilidades fundamentales para el diseño de herramientas de cribaje y la intervención tanto con niños como con adultos. Jordan et al. (2010) encontraron que el sentido numérico era un fuerte predictor del logro en matemáticas tanto en primer curso como en tercer curso de educación primaria. Otros estudios apoyan los resultados obtenidos por estos trabajos, por haber encontrado que las habilidades que conforman el sentido numérico juegan un papel importante en el desarrollo y adquisición de habilidades matemáticas o aritméticas más complejas (De Smedt, Verschaffel, & Ghesquière, 2009; Holloway & Ansari, 2009; Piazza, 2010).

La principal aportación de la presente investigación ha sido la de contrastar un modelo basado en la teoría, donde hemos podido constatar que el sentido numérico influye en más de una medida observada y que da cuenta de las correlaciones entre estas medidas observadas (i.e., comparación numérica, operaciones de dos dígitos, secuencias numéricas, operaciones de un dígito y valor de posición). Estas variables observadas están intercorrelacionadas porque comparten un factor causal común, esto es, se encuentran influenciadas por el mismo constructo subyacente: el sentido numérico.

Los mismos resultados fueron encontrados por Jiménez y De León (2017) para primer curso de educación primaria, lo que sugiere que las habilidades básicas subyacentes al sentido numérico se mantienen estables desde primer hasta tercer

curso de educación primaria. No obstante, se ha observado que en primer curso el factor latente *sentido numérico* influye principalmente en los indicadores observables *operaciones de un dígito* y *operaciones de dos dígitos* (Jiménez & De León, 2017); mientras que en tercer curso el factor latente *sentido numérico* influye principalmente en los indicadores observables *series numéricas*, *operaciones de un dígito* y *valor de posición*.

Una de las posibles limitaciones del presente estudio está relacionada con el conjunto de tareas que han sido seleccionadas, ya que no podemos descartar que otras tareas, por ejemplo la resolución de problemas verbales (ver Jordan et al., 2010; Sisco-Taylor et al., 2015), puedan ser igualmente válidas para la medición del sentido numérico.

Se encontró una covarianza de error estable en los tres momentos de medida (noviembre, febrero y mayo) entre las tareas OUD y ODD. Es posible que en estas edades una característica compartida por esos indicadores, y no explicada por la variable latente, esté relacionada con la capacidad para inhibir respuestas competitivas y seleccionar la respuesta más adecuada cuando se trata de resolver una operación aritmética (LeFevre et al., 2013; Lemaire & Lecacheur, 2011; Lucangeli et al., 2003). En este sentido, Lemaire y Lecacheur (2011) encontraron que las funciones ejecutivas y la edad se correlacionaban de manera significativa con la capacidad de la que disponen los niños para discriminar, entre distintas estrategias, aquella que es más adecuada para resolver una determinada tarea aritmética. En la misma línea, en un estudio anterior realizado por Fuchs et al. (2006) se encontró que la atención correlacionaba tanto con las operaciones numéricas básicas (operaciones de un dígito) como con aquellas más complejas (operaciones de dos dígitos o más).

Concluimos que el IPAM presenta una adecuada validez de constructo ya que permite captar

la capacidad de sentido numérico en estudiantes de tercer curso de educación primaria. A la vista de la revisión de estudios, se pone de manifiesto que esta capacidad estaría en la base de las dificultades que pueden experimentar algunos niños durante los primeros cursos de educación formal en el área de las matemáticas. Líneas futuras de investigación han de centrarse en analizar la sensibilidad y especificidad del IPAM para la detección de estudiantes en riesgo de presentar dificultades específicas de aprendizaje en matemáticas, así como su validez a la hora de evaluar el progreso en el aprendizaje.

Referencias

- Andrews, P., & Sayers, J. (2015). Identifying opportunities for grade one children to acquire foundational number sense: Developing a framework for cross cultural classroom analyses. *Early Childhood Education Journal*, 43(4), 257-267. doi: 10.1007/s10643-014-0653-6
- Cragg, L., & Gilmore, C. (2014). Skills underlying mathematics: The role of executive function in the development of mathematics proficiency. *Trends in Neuroscience and Education*, 3(2), 63-68. doi: 10.1016/j.tine.2013.12.001
- Dehaene, S. (2009). Origins of mathematical intuitions. The case of arithmetic. *Annals of the New York Academy of Sciences*, 1156(1), 232-259. doi: 10.1111/j.1749-6632.2009.04469.x
- De Smedt, B., Verschaffel, L., & Ghesquière, P. (2009). The predictive value of numerical magnitude comparison for individual differences in mathematics achievement. *Journal of Experimental Child Psychology*, 103(4), 469-479. doi: 10.1016/j.jecp.2009.01.010
- Dyson, N. I., Jordan, N. C., & Glutting, J. (2011). A number sense intervention for low-income kindergartners at risk for mathematics difficulties. *Journal of Learning Disabilities*, 46(2), 166-181. doi:

- 10.1177/0022219411410233
- Foegen, A., Jiban, C., & Deno, S. (2007). Progress monitoring measures in mathematics. A review of the literature. *The Journal of Special Education, 41*(2), 121-139. doi: 10.1177/00224669070410020101
- Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research 18*, 39-50. doi: 10.2307/3151312
- Fuchs, L. S., Fuchs, D., Compton, D. L., Powel, S. R., Seethaler, P. M., Capizzi, A. M. ... Fletcher, J. M. (2006). The cognitive correlates of third-grade skill in arithmetic, algorithmic computation, and arithmetic word problems. *Journal of Educational Psychology, 98*(1), 29-43. doi: 10.1037/0022-0663.98.1.29
- Geary, D. C. (2013). Early foundations for mathematics learning and their relations to learning disabilities. *Current Directions in Psychological Science, 22*(1), 23-27. doi: 10.1177/0963721412469398
- Gersten, R., Clarke, B., Jordan, N. C., Newman-Gonchar, R., Haymond, K., & Wilkins, C. (2012). Universal screening in mathematics for the primary grades: Beginnings of a research base. *Council for Exceptional Children, 78*(4), 423-445. doi: 10.1177/001440291207800403
- Hernández, J. A., & Betancort, M. (2016). ULLRToolbox. Disponible en <https://sites.google.com/site/ullrtoolbox>
- Hassinger-Das, B., Jordan, N. C., Glutting, J., Irwin, C., & Dyson, N. (2014). Domain-general mediators of the relation between kindergarten number sense and first-grade mathematics achievement. *Journal of Experimental Child Psychology, 118*, 78-92. doi: 10.1016/j.jecp.2013.09.008
- Holloway, I. D., & Ansari, D. (2009). Mapping numerical magnitudes onto symbols: The numerical distance effect and individual differences in children's mathematics achievement. *Journal of Experimental Child Psychology, 103*(1), 17-29. doi: 10.1016/j.jecp.2008.04.001
- Jiménez, J. E., & De León, S. (2016). *Indicadores de Progreso de Aprendizaje en Matemáticas (IPAM)*. Universidad de La Laguna. Manuscrito sin publicar.
- Jiménez, J. E., & De León, S. (2017). Análisis factorial confirmatorio de Indicadores de Progreso de Aprendizaje en Matemáticas (IPAM) en escolares de primer curso de primaria. *European Journal of Investigation in Health, Psychology and Education, 7*, 31-45. Recuperado de <https://formacionasunivep.com/ejihpe/index.php/ejihpe>
- Jitendra, A. K., Dupuis, D. N., & Zaslofsky, A. F. (2014). Curriculum-based measurement and standards-based mathematics: Monitoring the arithmetic word problem-solving performance of third-grade students at risk for mathematics difficulties. *Learning Disability Quarterly, 37*(4), 241-251. doi: 10.1177/0731948713516766
- Jordan, N. C., Gutting, J., & Ramineni, C. (2010). The importance of number sense to mathematics achievement in first and third grades. *Learning and Individual Differences, 20*, 82-88. doi: 10.1016/j.lindif.2009.07.004
- Jordan, N. C., Kaplan, D., Ramineni, C., & Locuniak, M. N. (2009). Early math matters: Kindergarten number competence and later mathematics outcomes. *Developmental Psychology, 45*(3), 850-867. doi: 10.1037/a0014939
- Jordan, N. C., & Levine, S. C. (2009). Socioeconomic variation, number competence, and mathematics learning difficulties in young children. *Developmental Disabilities Research Reviews, 15*(1), 60-68. doi: 10.1002/ddr.46
- Jöreskog, K., & Sörbom, D. (1996-2001). *LISREL 8: User's Reference Guide*. Illinois: Scientific Software International, Lincolnwood.
- Kim, D., Shin, J., & Lee, K. (2013). Exploring latent class based on growth rates in number sense ability. *Asia Pacific Education Review, 14*(3), 445-453, doi: 10.1007/s12564-013-9274-9
- Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2^a ed.). Nueva York, NY: Guilford Press.

- Kolkman, M. E., Kroesbergen, E. H., & Leseman, P. P. M. (2013). Early numerical development and the role of non-symbolic and symbolic skills. *Learning and Instruction, 25*, 95-103. doi: 10.1016/j.learninstruc.2012.12.001
- LeFevre, J., Berrigan, L., Vendetti, C., Kamawar, D., Bisanz, J., Skwarchuck, S. ... Smith-Chant, B. (2013). The role of executive attention in the acquisition of mathematical skills for children in grades 2 through 4. *Journal of Experimental Child Psychology, 114*(2), 243-261. doi: 10.1016/j.jecp.2012.10.005
- Lemaire, P., & Lecacheur, M. (2011). Age-related changes in children's executive functions and strategy selection: A study in computational estimation. *Cognitive Development, 26*, 282-294. doi: 10.1016/j.cogdev.2011.01.002
- Lembke, E., & Foegen, A. (2009). Identifying early numeracy indicators for kindergarten and first-grade students. *Learning Disabilities Research & Practice, 24*(1), 12-20. doi: 10.1111/j.1540-5826.2008.01273.x
- Libertus, M. E., Fiegenson, L., & Halberda, J. (2011). Preschool acuity of the approximate number system correlates with school math ability. *Developmental Science, 14*(6), 1292-1300. doi: 10.1111/j.1467-7687.2011.01080.x
- Libertus, M. E., Fiegenson, L., & Halberda, J. (2013). Is approximate number precision a stable predictor of math ability? *Learning and Individual Differences, 25*, 126-133. doi: 10.1016/j.lindif.2013.02.001
- Lucangeli, D., Tressoldi, P. E., Bendotti, M., Bonanomi, M., & Siegel, L. S. (2003). Effective strategies for mental and written arithmetic calculation from the third to the fifth grade. *Educational Psychology, 23*(5), 507-520. doi: 10.1080/0144341032000123769
- Lyons, I. M., & Beilock, S. L. (2011). Numerical ordering ability mediates the relation between number-sense and arithmetic competence. *Cognition, 121*(2), 256-261. doi: 10.1016/j.cognition.2011.07.009
- Lyons, I. M., Price, G. R., Vaessen, A., Blomert, L., & Ansari, D. (2014). Numerical predictors of arithmetic success in grades 1-6. *Developmental Science, 17*(5), 714-726. doi: 10.1111/desc.12152
- Namkung, J. M., & Fuchs, L. S. (2016). Cognitive predictors of calculations and number line estimation with whole numbers and fractions among at-risk students. *Journal of Educational Psychology, 108*(2), 214-228. doi: 10.1037/edu0000055
- National Council of Teachers of Mathematics (2000). *Principles and standards for school mathematics*. Reston, VA: NCATE.
- National Mathematics Advisory Panel (2008). *Foundations for Success: The Final Report of the National Mathematics Advisory Panel*. Washington, DC: Department of Education.
- Park, J., & Brannon, E. M. (2014). Improving arithmetic performance with number sense training: An investigation of underlying mechanism. *Cognition, 133*(1), 188-200. doi: 10.1016/j.cognition.2014.06.011
- Peng, P., Namkung, J. M., Fuchs, D., Fuchs, L. S., Patton, S., Yen, L. ... Hamlett, C. (2016). A longitudinal study on predictors of early calculation development among young children at risk for learning difficulties. *Journal of Experimental Child Psychology, 152*, 221-241. doi: 10.1016/j.jecp.2016.07.017
- Piazza, M. (2010). Neurocognitive start-up tools for symbolic number representations. *Trends in Cognitive Sciences, 14*(12), 542-551. doi: 10.1016/j.tics.2010.09.008
- R Core Team (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Recuperado de <http://www.R-project.org>
- Sasanguie, D., Göbel, S. M., Moll, K., Smets, K., & Reynvoet, B. (2013). Approximate number sense, symbolic number processing, or number-space mappings: What underlies mathematics achievement? *Journal of Experimental Child Psychology, 114*(3), 418-431. doi: 10.1016/j.jecp.2012.10.012
- Sayers, J., & Andrews, P. (2015). Foundational number sense: Summarising the development of an analytical framework. En K. Krainer & N. Vondrová (Eds.), *Ninth Congress of the European Society for Research*

in *Mathematics Education (CERME9)* (pp. 361-337). Praga: Charles University in Prague, Faculty of Education.

- Schneider, M., Beeres, K., Coban, L., Merz, S., Schmidt, S. S., Stricker, J. ... De Smedt, B. (2016). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: A meta-analysis. *Developmental Science*, 20(3), 1-16. doi: 10.1111/desc.12372
- Siegler, R., & Araya, R. (2005). A computational model of conscious and unconscious strategy discovery. *Advances in Child Development and Behavior*, 33, 1-42. doi: 10.1016/S0065-2407(05)80003-5
- Sisco-Taylor, D., Fung, W., & Swanson, H. L. (2015). Do curriculum-based measures predict performance on word-problem-solving measures? *Assessment for Effective Intervention*, 40(3), 131-142. doi: 10.1177/1534508414556504
- Starr, A., Libertus, M. E., & Brannon, E. M. (2013). Number sense in infancy predicts mathematical abilities in childhood. *Proceedings of the National Academy of Science*, 110(45), 18116-18120. doi: 10.1073/pnas.1302751110
- Toll, S. W. M., Kroesbergen, E. H., & Van Luit, J. E. H. (2016). Visual working memory and number sense: Testing the double deficit hypothesis in mathematics. *British Journal of Educational Psychology*, 86(3), 429-445. doi: 10.1111/bjep.12116
-

Propiedades Psicométricas de la BPNES Adaptada a la Educación Física en Argentina, España y Colombia

Psychometric properties of the BPNES adapted to Physical Education in Argentina, Spain and Colombia

Evelia Franco *¹, Javier Coterón², Valeria Gómez³, Jorge Brito⁴

1 - Universidad Pontificia de Comillas, España.

2 - Universidad Politécnica de Madrid, España.

3 - Universidad de Flores, Argentina.

4 - Universidad de Cuenca, Ecuador.

Introducción
Método
Instrumentos
Resultados
Discusión
Referencias

Recibido: 25/07/2017 Revisado: 04/09/2017 Aceptado: 08/09/2017

Resumen

El objetivo del presente trabajo fue examinar la medida en que las propiedades psicométricas de la Escala de Satisfacción de las Necesidades Psicológicas Básicas (BPNES) resultan estables en muestras de estudiantes de educación física de Argentina, España y Colombia. Participaron 1153 estudiantes de educación física (44.6% varones) de edades comprendidas entre 13 y 16 años ($M = 14.68$; $DE = 1.09$). Los análisis factoriales en muestras independientes revelaron adecuadas propiedades psicométricas del instrumento en los tres países. Los análisis multigrupo posteriores demostraron que la estructura factorial, las cargas factoriales y los interceptos eran invariantes entre los grupos. Sin embargo, se encontraron diferencias entre las varianzas residuales. Se discuten los resultados con base en los hallazgos relativos a la consistencia interna de las dimensiones del instrumento y las diferencias observadas entre las tres muestras. Los hallazgos de este trabajo respaldan posteriores comparaciones interculturales mediante el uso de la BPNES.

Palabras clave: *análisis factorial, necesidades psicológicas básicas, invarianza, motivación*

Abstract

The main goal of this study was to examine the extent to which psychometric properties from the Basic Psychological Needs in Exercise Scale (BPNES) are invariant across samples of Argentina, Spanish and Colombian physical education students. The questionnaire was distributed to 1153 students (44.6% male) from ages 13 to 16 ($M = 14.68$; $SD = 1.09$). Confirmatory factor analyses carried out in independent samples revealed suitable psychometric properties in the three samples. Multi-sample confirmatory factor analyses showed that factor structure, factor loadings and intercepts can be considered invariant across the groups. However, residual variances were found to be different across some of the samples. Findings are discussed in terms of internal consistency and differences observed between the three samples. Findings support valid cross-cultural comparisons using BPNES.

Keywords: *factor analysis, basic psychological needs, invariance, motivation*

*Correspondencia a: Evelia Franco Álvarez, e.francoalvarez@gmail.com

Cómo citar este artículo: Franco, E., Coterón, J., Gómez, V., & Brito, J. (2017). Propiedades psicométricas de la BPNES adaptada a la educación física en Argentina, España y Colombia. *Revista Evaluar*, 17(2), 97-112. Recuperado de <https://revistas.unc.edu.ar/index.php/revaluar>

Introducción

A pesar de los innegables beneficios físicos y psicológicos que una vida activa tiene sobre la salud, durante los últimos años se observa un importante descenso del nivel de actividad física (AF) realizada durante la adolescencia (Meneses-Montero & Ruiz-Juan, 2017; Verloigne et al., 2012). Diferentes estudios indican que la educación física (EF) contribuye significativamente a los niveles de práctica de AF en niños y adolescentes (Aibar-Solana et al., 2015); esto es debido a su presencia curricular obligatoria en etapas educativas tempranas de todos los países, lo que facilita un entorno en el que los jóvenes puedan mantener un contacto cotidiano con el deporte y la AF.

La motivación demostrada por los alumnos en el contexto de EF es uno de los factores que más frecuentemente se han relacionado en los últimos años con la práctica deportiva en adolescentes, y es destacable la influencia de la disposición motivacional en los niveles de AF (Standage, Sebire, & Loney, 2008).

La teoría de la autodeterminación y el rol de las necesidades psicológicas básicas

La teoría de la autodeterminación (TAD; Deci & Ryan, 1985) supone un marco teórico ampliamente utilizado en el abordaje de estudios relativos a la motivación en EF (Franco, Coterón, & Gómez, 2017; Van Aart, Hartman, Elferink-Gemser, Mombarg, & Visscher, 2017).

Según esta teoría, las personas somos organismos activos con una tendencia innata hacia el crecimiento y el desarrollo psicológico; esta se centra en el análisis del grado en que las conductas humanas son autodeterminadas, es decir: la medida en que realizamos las acciones de for-

ma voluntaria al nivel más alto de reflexión y nos comprometemos con las mismas con un sentido de elección (Deci & Ryan, 1985). A través de una de las seis mini-teorías que componen este marco teórico, la teoría de las necesidades básicas, se postula que dicho comportamiento es motivado y regulado por tres necesidades psicológicas primarias y universales: autonomía, competencia y relación con los demás; y la satisfacción de las mismas se considera determinante del grado y tipo de motivación que los individuos presenten.

En el marco de la clase de EF, la necesidad de *autonomía* considera los esfuerzos de los estudiantes por sentirse el origen de sus acciones y poder determinar su propio comportamiento. La necesidad de *competencia* se basa en tratar de controlar el resultado y experimentar eficacia en su participación durante las actividades propuestas. Por último, la necesidad de *relación con los demás* hace referencia a la necesidad de relacionarse y preocuparse por otros y se define mediante dos dimensiones: sentirse aceptado e intimar con los demás (Ryan, 1991).

Las investigaciones indican que cada una de ellas juega un papel importante para el desarrollo y la experiencia óptima y que la satisfacción de las tres necesidades influye positivamente sobre la motivación intrínseca; mientras que la frustración de las mismas se asocia con una menor motivación intrínseca y una mayor motivación extrínseca y desmotivación (Deci & Ryan, 2000).

Siguiendo los postulados de esta teoría, se han llevado a cabo numerosos estudios relacionados con la AF que demostraron la influencia que la satisfacción de estas necesidades en clase de EF tiene sobre variables relacionadas con la práctica de AF, tanto directamente como indirectamente, a través de la mejora de la motivación autodeterminada (Aibar-Solana et al., 2015; Chatzisarantis & Hagger, 2009).

Medición de las necesidades psicológicas básicas en contextos de AF

Con base en lo anteriormente expuesto, se han creado numerosos instrumentos para medir la satisfacción de las necesidades psicológicas básicas (NPB) tales como la Basic Need Satisfaction Scale para contextos de EF (BNS-PE; Ntoumanis, 2005), una adaptación de la escala diseñada por Deci et al. (2001) para contextos laborales, la Psychological Need Satisfaction in Exercise Scale (PNSE; Wilson, Rogers, Rodgers, & Wild, 2006), o la Escala de Mediadores Motivacionales en el Deporte (EMMD; González-Cutre et al., 2007).

Una de las herramientas más utilizadas en la medida de la satisfacción de las NPB y sus relaciones con la AF es la Basic Psychological Needs in Exercise Scale (BPNES; Vlachopoulos & Michailidou, 2006). Esta herramienta fue creada y validada para la medición de la satisfacción de las NPB en contextos de ejercicio a partir de los cuestionarios previos de la TAD (Deci & Ryan, 2001). Se crearon 31 ítems iniciales a partir de los cuales tres expertos del ámbito trabajaron hasta concluir la versión final compuesta por 12 ítems (cuatro para cada una de las necesidades). Los ítems se responden en una escala Likert de 1 (*Totalmente en desacuerdo*) a 5 (*Totalmente de acuerdo*). La BPNES ha sido validada en España tanto en el ámbito del ejercicio físico (Sánchez & Núñez, 2007) como en las clases de EF (Moreno-Murcia, González-Cutre-Coll, Chillón-Garzón, & Parra-Rojas, 2008).

La adecuación de las propiedades psicométricas de la BPNES en diferentes contextos (Cid et al., 2016; Evelein, Korthagen, & Brekelmans, 2008; Moreno-Murcia, Martínez-Galindo, Moreno-Pérez, Marcos, & Borges, 2012; Moutao, Serra-Majem, Alves, Leitao, & Vlachopoulos, 2012; Sánchez & Núñez, 2007; Vlachopoulos, 2008; Vlachopoulos et al., 2013; Vlachopoulos,

Ntoumanis, & Smith, 2010) y especialmente el éxito con el que ha sido utilizada en EF (Cuevas, García-López, & Contreras, 2015; García-González, Aibar, Sevil, Almolda-Tomás, & Julián-Clemente, 2015; Gómez-Rijo, Hernández-Moreno, Martínez-Herráez, & Gámez, 2014; Méndez-Giménez, Fernández-Río, & Cecchini, 2013; Sevil, Abós, Julián-Clemente, Murillo, & García-González, 2015), junto al asequible número de ítems y la redacción clara y sencilla de los mismos, emergen como los principales motivos para trabajar en la adaptación y validación de esta herramienta con el fin de ampliar su uso a diferentes contextos geográficos.

Generalización intercultural

En tanto que la cultura puede ser una importante influencia sobre diversas variables psicológicas, se hacen necesarios estudios que incorporen análisis interculturales con el fin de evitar la propagación de teorías que puedan no ser generalizables (Duda & Allison, 1990; Duda & Hayashi, 1998). Así, en el marco de la TAD, la literatura muestra cómo, si bien existen patrones de comportamiento similares, en ocasiones los constructos principales de la teoría se han relacionado de un modo ligeramente diferente en función del contexto cultural (Cantú-Berrueto, Castillo, López-Walle, Tristán, & Balaguer, 2016; Pulido, Sánchez-Oliva, González-Ponce, Amado, & Chamorro, 2017; Van Aart et al., 2017).

En los últimos años, está emergiendo en Latinoamérica una interesante línea de trabajo sobre la motivación en adolescentes, tanto en contextos deportivos (López-Walle, Balaguer, Castillo, & Tristán, 2011; López-Walle, Tomás, Castillo, Tristán, & Balaguer, 2011; Reyes, Raimundi, & Gómez-Correa, 2012) como de EF (Franco, Coterón, Gómez, & De Franza, 2017; Franco,

Coterón, Martínez, & Brito, 2017; Zamarripa, Castillo, Tomás, & López-Walle, 2016). Estos trabajos han abordado el estudio de las características motivacionales de estos practicantes de actividad física y la relación entre dichas características y diversos patrones comportamentales; y han encontrado, independientemente del país en el que se realizara el estudio, una relación positiva entre la motivación y patrones más adaptativos en lo que a práctica de actividad física se refiere.

Considerando esta realidad, es necesario tener en cuenta que el uso de un instrumento de medida en un nuevo contexto cultural supone algo más que la mera traducción, administración de cuestionarios y comparación de resultados. En la investigación psicológica resulta de interés la realización de tests de invarianza de instrumentos de medida antes de llevar a cabo análisis de diferencias interculturales (Byrne, 2004), y se deben comprobar las propiedades psicométricas de una escala en los diferentes contextos en que se va a utilizar antes de interpretar las diferencias encontradas (Cheung & Rensvold, 2002) para garantizar que los ítems de los cuestionarios adquieren el mismo sentido para sujetos de diferentes grupos. En resumen, sin explorar la estructura de los constructos y la adecuación de los ítems para comprobar que en diferentes contextos mantienen el mismo significado, no tendría sentido emprender trabajos de comparación entre países (He & Van de Vijver, 2012).

Como se ha mencionado anteriormente, la BPNES adaptada a la EF ha mostrado una adecuada validez y confiabilidad en diferentes países iberoamericanos tales como España (Moreno-Murcia et al., 2008), Brasil (Moutao et al., 2012), México (Zamarripa et al., 2016) y Portugal (Cid et al., 2016). A pesar de la proliferación de trabajos de la temática desarrollados en países latinoamericanos hispanohablantes, no se han encontrado estudios que evalúen las propiedades

psicométricas de este instrumento en Argentina y Colombia. Por lo tanto, el objetivo principal del presente estudio fue analizar las propiedades psicométricas del instrumento con muestras de estos países. Considerando que la BPNES había sido previamente validada en estudiantes de EF españoles, y que la mayor parte de los estudios sobre satisfacción de NPB en contextos hispanohablantes han sido desarrollados en España, se decidió incluir información de adolescentes de este país para, en un segundo paso, evaluar la invarianza de medida del instrumento en estudiantes de Argentina, España y Colombia. Como segundo objetivo, nos planteamos el análisis de posibles diferencias en la satisfacción de las NPB entre los distintos países.

Método

Participantes

Se trabajó con un muestreo no probabilístico incidental considerando como criterios de inclusión que los alumnos asistieran a centros educativos ordinarios, participaran en clase de EF con normalidad y tuvieran entre 13 y 16 años de edad. Las respuestas de la BPNES se obtuvieron de 1153 estudiantes de EF (44.6% varones) de edades comprendidas entre 13 y 16 años ($M = 14.68$; $DE = 1.09$) de colegios públicos y privados de Argentina ($n = 224$), España ($n = 450$), y Colombia ($n = 479$). En la Tabla 1 se presenta la distribución de la muestra por países, así como los estadísticos descriptivos de género y edad en los participantes de cada país.

Tabla 1

Distribución de la muestra por país y género, y estadísticos descriptivos de la edad en cada país.

	N total	Hombres	Mujeres	Edad	
				Rango	M (DE)
España	450	205	236	13 - 16	14.24 (1.05)
Argentina	224	82	142	13 - 16	14.99 (.95)

Instrumentos

Se utilizó la versión española adaptada a la EF (Moreno-Murcia et al., 2008) de la BPNES (Vlachopoulos & Michailidou, 2006). La versión utilizada estaba encabezada con el enunciado *En mis clases de EF...* y constaba de 12 ítems, cuatro para medir cada una de las dimensiones (satisfacción de la necesidad de *autonomía*, satisfacción de la necesidad de *competencia* y satisfacción de la necesidad de *relación con los demás*). En la Tabla 2 se presentan los ítems que componen la escala. Dichos ítems se responden en una escala Likert de 1 (*Totalmente en desacuerdo*) a 5 (*Totalmente de acuerdo*). Diferentes estudios han proporcionado evidencias de la validez y confiabilidad del instrumento en diferentes contextos (Cid et al., 2016; Moreno-Murcia et al., 2012; Moutao et al., 2012; Sánchez & Núñez, 2007). Los valores de alfa de Cronbach para la satisfacción de *autonomía*, *competencia* y *relación con los demás* en este estudio fueron respectivamente: .74, .65 y .80 en la muestra argentina; .76, .71 y .84 en la muestra española, y .77, .71 y .79 en la muestra colombiana.

Diseño y procedimiento

Se trata de un estudio transversal descriptivo. Los datos fueron recogidos de muestras no probabilísticas de estudiantes de EF de los tres países participantes. Se estableció contacto con los equipos directivos y/o responsables del de-

partamento de EF de los diferentes centros de enseñanza para informarles de los objetivos del estudio y pedirles su colaboración. Una vez obtenidas las autorizaciones pertinentes para que los alumnos pudieran participar en el estudio, los cuestionarios fueron administrados durante la clase de EF por un miembro del equipo investigador previamente formado.

Esta persona proporcionó una breve explicación del objetivo del estudio, informó a los participantes sobre la forma de completar el cuestionario e insistió acerca del anonimato de las respuestas solicitando sinceridad a los participantes. El tiempo requerido para completar el cuestionario fue de aproximadamente 15 minutos, variando ligeramente según la edad de los estudiantes. La recolección de datos siguió las indicaciones éticas de la Asociación de Psicología Americana (APA, 2002).

Análisis de datos

Se examinaron la consistencia interna, con base en el índice de confiabilidad compuesta para cada subescala de la BPNES y el coeficiente alfa de Cronbach; y la estructura factorial a partir de análisis factoriales confirmatorios.

Se consideraron diferentes índices de bondad de ajuste: ratio entre chi-cuadrada y grados de libertad (χ^2/gl), RMSEA (*Root Mean Square Error of Approximation*) e índice incremental CFI (*Comparative Fit Index*). Estos índices de bondad de ajuste son considerados aceptables cuan-

Tabla 2

Ítems que componen la versión del BPNES administrada (Moreno-Murcia et al., 2008).

Dimensiones	Ítems
Autonomía	1. Los ejercicios que realizo se ajustan a mis intereses.
	4. La forma de realizar los ejercicios coincide perfectamente con la forma en que yo quiero hacerlos.
	7. La forma de realizar los ejercicios responde a mis deseos.
	10. Tengo la oportunidad de elegir cómo realizar los ejercicios.
Competencia	2. Siento que he tenido una gran progresión con respecto al objetivo final que me he propuesto.
	5. Realizo los ejercicios eficazmente.
	8. El ejercicio es una actividad que hago muy bien.
	11. Creo que puedo cumplir con las exigencias de la clase.
Relación con los demás	3. Me siento muy cómodo/a cuando hago ejercicio con los/as demás compañeros/as.
	6. Me relaciono de forma muy amistosa con el resto de compañeros.
	9. Siento que me puedo comunicar abiertamente con mis compañeros/as.
	12. Me siento muy cómodo/a con los/as compañeros/as.

do χ^2/gl presenta valores inferiores a 5 (Bentler, 1989), RMSEA muestra valores por debajo de 0.8 (Browne & Cudeck, 1993); y el CFI está por encima de 0.9 (Hu & Bentler, 1995; Marsh, Hau, & Wen, 2004).

Se utilizó el método de estimación de máxima verosimilitud en base a la matriz de covarianzas junto al procedimiento de *bootstrapping*, lo que permitió asumir que los datos eran robustos ante la falta de normalidad (Byrne, 2004).

La invarianza se evaluó basándose en un análisis factorial confirmatorio multigrupo. La comparación de modelos fue realizada según la ΔCFI y la $\Delta RMSEA$. Un valor de .01 o menor en la ΔCFI y de .015 o menor en la $\Delta RMSEA$ indicarían ausencia de diferencia entre los modelos (Chen, 2007; Cheung & Rensvold, 2002).

Por último, se llevó a cabo un ANOVA para analizar diferencias entre grupos. Se realizó

el test de Scheffé para comparar las diferencias entre las puntuaciones de los participantes de los diferentes países.

Se utilizaron los paquetes estadísticos SPSS (IBM, 2016) y AMOS (Arbuckle, 2014) para el análisis de los datos.

Resultados

El coeficiente de curtosis multivariante de Mardia (50.30) indicó la ausencia de normalidad multivariante (Raykov & Marcoulides, 2006).

El contraste de la equivalencia factorial comenzó con un análisis preliminar en el que se examinó por separado la bondad de ajuste de la estructura de la BPNES en las muestras de estudiantes de los tres países. Como se muestra en la Tabla 3, el coeficiente entre el chi cuadrada y los grados

Tabla 3

Índices de bondad de ajuste del modelo factorial confirmatorio en los tres países.

	χ^2	<i>p</i>	<i>df</i>	χ^2/df	CFI	RMSEA	RMSEA 90% IC	Cargas factoriales estand.	Fiabilidad Compuesta		
									A	C	R
Argentina	87.82	.001	51	1.72	.953	.057	.036-.077	.49-.81	.74	.67	.80
España	151.87	.001	51	2.98	.944	.066	.054-.079	.50-.83	.74	.77	.86
Colombia	197.30	.001	51	3.87	.919	.077	.066-.089	.48-.81	.75	.77	.83

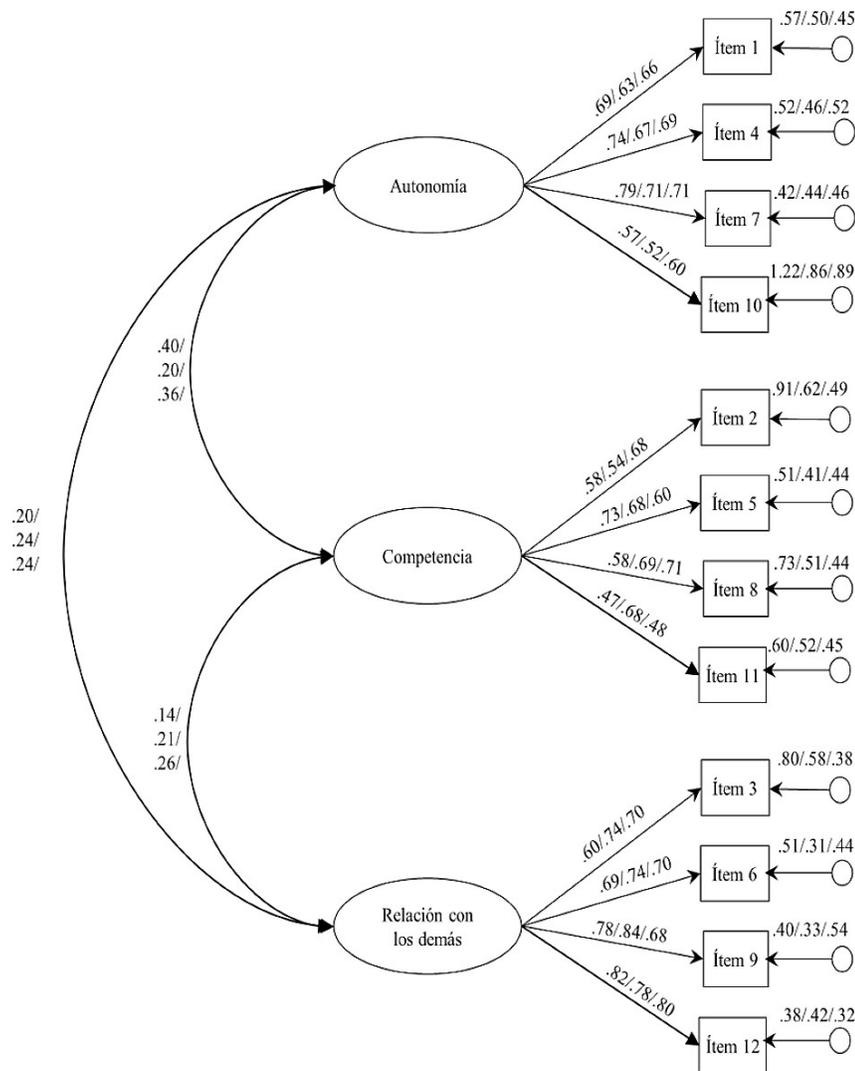
de libertad se sitúa por debajo de 5 en todas las muestras. Además, el CFI está por encima de .90 en todos los casos mientras que el RMSEA se encuentra por debajo de .8 en los tres grupos. Según las recomendaciones de [Herrero \(2010\)](#), si un modelo presenta un buen ajuste a través del CFI y del RMSEA conjuntamente, es muy poco probable que el modelo no sea adecuado a los datos.

Tanto los índices de confiabilidad compuesta como el coeficiente alfa de Cronbach resultaron adecuados en líneas generales. Cabe mencionar que los datos de la muestra argentina revelaron una consistencia débil en la dimensión de satisfacción de necesidad de *competencia* situándose el valor alfa y la confiabilidad compuesta por debajo de los índices recomendados, si bien se decidió interpretar favorablemente los resultados al tratarse de valores cercanos a los establecidos como adecuados.

Las cargas factoriales de los indicadores en sus respectivas variables latentes estuvieron por encima de .50 con excepción del cuarto indicador de competencia en la muestra argentina (Figura 1). El bajo peso de regresión de este indicador sobre su variable latente podría explicar los bajos valores observados en el valor alfa y la confiabilidad compuesta de la dimensión de *competencia*. Sin embargo, se decidió no suprimir el ítem para evitar la pérdida de información válida para el modelo evaluado ([Chin, 1998](#)).

A continuación, se realizaron los análisis multigrupo creando nuevos modelos anidados ([Gregorich, 2006](#)) cuyos índices de ajuste se muestran en la Tabla 4. El Modelo 1 (sin restricciones) examinó la invarianza estructural de la BPNES en los tres grupos mostrando adecuados índices de ajuste. Este modelo se consideró como referencia para la subsiguiente anidación de restricciones. El Modelo 2 (restricción de las cargas factoriales) también resultó invariante ($\Delta CFI < .01$), lo que indica la ausencia de diferencias entre las cargas factoriales de las muestras de Argentina y del resto de los países. El Modelo 3 (en el que se agrega la equivalencia de interceptos) tampoco excede el valor criterio de .01 para ΔCFI pudiendo aceptarse la hipótesis de invarianza de los interceptos. Los índices de ajuste del Modelo 4 (en el que se agrega la equivalencia de los residuos de los ítems) revelan la ausencia de invarianza residual de las muestras de Argentina con España ($\Delta CFI = .025$) y Colombia ($\Delta CFI = .31$).

La Tabla 5 muestra la comparación entre las puntuaciones de la muestra argentina y el resto. Los estudiantes argentinos mostraron puntuaciones menores que los participantes españoles en la satisfacción de la necesidad de *competencia* ($M_{\text{Argentina}} = 3.69$; $M_{\text{España}} = 3.83$) y de *relación con los demás* ($M_{\text{Argentina}} = 3.85$; $M_{\text{España}} = 4.15$), no se hallaron diferencias en la satisfacción de *autonomía* ($M_{\text{Argentina}} = 3.15$; $M_{\text{España}} = 3.20$) entre

**Figura 1**

Parámetros estandarizados (covarianzas entre factores, pesos de regresión y errores de medida) del análisis factorial confirmatorio del BPNES en Argentina, España y Colombia.

Nota. Los parámetros de los diferentes países se presentan separados por / (Argentina/España/Colombia).

estos dos países. Por otra parte, los adolescentes argentinos mostraron menores puntuaciones que los colombianos en la satisfacción de *autonomía* ($M_{\text{Argentina}} = 3.15$; $M_{\text{Colombia}} = 3.48$), de *competencia*, ($M_{\text{Argentina}} = 3.69$; $M_{\text{Colombia}} = 3.89$) y de *relación con los demás* ($M_{\text{Argentina}} = 3.85$; $M_{\text{Colombia}} = 3.99$).

Discusión

El objetivo del presente estudio fue analizar las propiedades psicométricas y la invarianza de medida de la BPNES en muestras de estudiantes de EF de países diferentes que comparten una misma lengua (Argentina, España y Colombia).

Tabla 4

Índices de bondad de ajuste de cada uno de los modelos evaluados en la invarianza factorial de la BPNES. Comparación entre los modelos anidados.

Modelo	χ^2	<i>gl</i>	CFI	Δ CFI	RMSEA	Δ RMSEA	RMSEA 90% IC
<i>Argentina - España</i>							
Modelo 1	239.71	102	.947	-	.045	-	.037-.052
Modelo 2	254.56	111	.944	.003	.044	.001	.037-.051
Modelo 3	269.42	117	.941	.003	.044	.000	.037-.051
Modelo 4	346.44	129	.916	.025	.050	.006	.044-.056
<i>Argentina - Colombia</i>							
Modelo 1	285.11	102	.930	-	.051	-	.044-.058
Modelo 2	293.87	111	.930	.000	.048	.003	.042-.055
Modelo 3	312.18	117	.925	.005	.049	.001	.042-.055
Modelo 4	404.18	129	.894	.031	.055	.006	.049-.061

Tabla 5

Diferencias entre la muestra argentina y el resto de los países en la satisfacción de las necesidades psicológicas básicas.

	Autonomía			Competencia			Relación con los demás		
	<i>t</i>	<i>p</i>	<i>d</i> Cohen	<i>t</i>	<i>p</i>	<i>d</i> Cohen	<i>t</i>	<i>p</i>	<i>d</i> Cohen
Argentina-España	-.81	.42	.05	-2.32	.020	.19	-4.28	.001	.35
Argentina-Colombia	-4.76	.001	.39	-3.29	.001	.27	-2.09	.038	.17

Esto es, analizar si “el instrumento de medida funciona exactamente de la misma manera y el constructo que está siendo medido tiene la misma estructura teórica para cada grupo de estudio” (Byrne, 2004; p. 272).

El análisis factorial confirmatorio del BPNES desarrollado con cada una de las muestras apoya la estructura del instrumento. De acuerdo con expertos en medición (Cheung & Rensvold, 2002; Gregorich, 2006), y en línea con trabajos previos que han evaluado la invarianza cultural del BPNES (Cid et al., 2016; Vlachopoulos et al., 2013) así como de otros instrumentos que miden variables motivacionales en entornos de AF

(López-Walle, Tomás, et al., 2011; Wang, Pyun, Kim, & Chatzisarantis, 2009), se examinó la invarianza estructural a través de un análisis factorial multigrupo.

Los análisis factoriales confirmatorios independientes llevados a cabo con cada una de las muestras utilizadas revelaron que el modelo de medida mostraba unos índices adecuados en los tres países. Las cargas factoriales estandarizadas de cada uno de los indicadores sobre sus factores son aceptables y, en la mayor parte de los casos, aceptadas (Barclay, Higgins, & Thompson, 1995; Chin, 1998). A pesar de que en términos generales estos pesos de regresión son adecuados, uno de

los indicadores de la dimensión de *competencia* muestra una carga factorial muy baja en el análisis factorial llevado a cabo en Argentina y Colombia (.47 y .48 respectivamente). Este indicador se corresponde con el ítem *En mis clases de EF pienso que puedo cumplir con las exigencias de la clase*. Si consideramos el resto de los ítems utilizados para medir esta dimensión (*Siento que he tenido una gran progresión con respecto al objetivo final que me he propuesto; Realizo los ejercicios eficazmente y El ejercicio es una actividad que hago muy bien*), se puede comprobar la diferencia existente entre el primer ítem y este último grupo en función del criterio de éxito utilizado para evaluar la competencia (Nicholls, 1984a). Mientras que en los ítems que revelaron una adecuada carga factorial se evidencia un criterio autorreferenciado de éxito (no es necesario que el estudiante se compare con alguien para evaluar su competencia), en el caso del ítem que mostró un bajo peso de regresión la valoración de la competencia se hace en base a un criterio normativo (el éxito del estudiante depende de lo que exija la clase).

Estos datos apuntan a que, si bien la muestra española parece percibir su competencia indistintamente de cuál sea el criterio de referencia, los estudiantes de Argentina y Colombia conceptualizan la competencia de forma diferente en función del criterio de éxito utilizado. Estudios anteriores que han analizado las percepciones de éxito desde el enfoque de la teoría de las metas de logro (Nicholls, 1984b) en países latinoamericanos y España (Franco, Coterón, & Martínez, en prensa; López-Walle, Tomás, et al., 2011) han encontrado que los adolescentes de países latinoamericanos puntuaban por encima de los españoles en la orientación al ego, adoptando en mayor medida un criterio normativo de percepción de competencia. Los resultados del presente trabajo, junto a los hallazgos de estos estudios anteriores, sugieren que el escaso peso de regresión del último

indicador de satisfacción de *competencia* podría mejorarse con la modificación de este ítem para que esté en línea con el criterio autorreferenciado presente en los otros tres indicadores.

De la misma manera, y probablemente a consecuencia de lo anteriormente expuesto, la confiabilidad compuesta del factor de *competencia* en la muestra argentina estuvo por debajo de los límites considerados aceptables (Fornell & Larcker, 1981). Este resultado refuerza la sugerencia anterior de desarrollar estudios futuros que estudien la confiabilidad del factor de *competencia* unificando el criterio de éxito en el que se basen los ítems utilizados como indicadores.

En cuanto al análisis multigrupo, los resultados del presente estudio mostraron evidencia de invarianza estructural en todas las muestras. Esto sugiere que los constructos de satisfacción de *autonomía, competencia y relación con los demás* se conceptualizaron de forma similar en los diferentes grupos, a pesar de la limitación encontrada con el cuarto indicador del factor de *competencia*. A pesar de que puede existir otro tipo de diferencias entre los países, el contexto cultural no afecta las percepciones de los estudiantes sobre los constructos medidos (Millsap & Kwok, 2004).

Los resultados también apoyaron la invarianza en las cargas factoriales, lo que significa que los participantes de Argentina, España y Colombia realizaron una interpretación similar de la BPNES y coincidieron en el número de constructos, así como en los ítems asociados a cada constructo (Cheung & Rensvold, 2002). Según Bollen (1989) la invarianza en las cargas factoriales supone un requisito imprescindible para las comparaciones entre grupos, por lo que los hallazgos del presente estudio pueden sostener futuras investigaciones que busquen comparar la satisfacción de las necesidades psicológicas básicas utilizando la BPNES en estos países.

La invarianza en los interceptos indicó que

las diferencias interculturales entre las variables latentes indicadas por los ítems son las mismas en los diferentes ítems (Cheung & Rensvold, 2002), lo que permite llevar a cabo comparaciones entre las muestras de los distintos países tanto de las variables latentes como de las variables observadas.

Los resultados no apoyaron la invarianza en las varianzas residuales de la muestra argentina con la española ni con la colombiana. Basándose en los hallazgos del presente estudio, las comparaciones de las varianzas observadas entre las respuestas obtenidas por los estudiantes argentinos y los españoles o los colombianos no se garantizan. Sin embargo, esto no supone un serio problema ya que las comparaciones de las varianzas no son útiles al comparar resultados de diferentes culturas.

El segundo objetivo del presente estudio fue analizar posibles diferencias entre estudiantes argentinos, españoles y colombianos. Se encontró que los adolescentes argentinos mostraron menor satisfacción de las necesidades de *competencia* y *relación con los demás* que los estudiantes españoles, así como puntuaciones inferiores que los adolescentes colombianos en las tres variables.

Existen estudios anteriores que han comparado variables motivacionales frecuentemente asociadas con la satisfacción de las NPB, como la motivación intrínseca o el *flow* disposicional, en estudiantes de EF de Argentina y España (Franco, Coterón, Gómez, Brito, & Martínez-González, 2017; Franco, Coterón, Gómez, & De Franza, 2017). Los resultados de estos estudios han señalado que los adolescentes argentinos tienden a mostrar puntuaciones ligeramente inferiores a las de estudiantes de otros países, lo que acarrea patrones comportamentales menos adaptativos en lo que a práctica de AF se refiere (Franco, Coterón, Gómez, Brito, et al., 2017).

Si bien no se han hallado estudios que hayan comparado la satisfacción de las NPB entre

estudiantes de EF de Argentina y otros países, sí existen en la literatura referencias en las que se ha medido la satisfacción de estas necesidades en Argentina (Franco, Coterón, & Gómez, 2017) y en España independientemente (Méndez-Giménez, et al., 2013). Los resultados de estos trabajos, en línea con los hallazgos del presente estudio, sugieren que si bien las puntuaciones en la variable de satisfacción de autonomía no difieren entre los adolescentes de ambos países, los adolescentes españoles muestran niveles más elevados en la satisfacción de *competencia* y de *relación con los demás*. En el trabajo realizado por Franco, Coterón y Gómez (2017) se comprobó la adecuación de un modelo predictivo según el cual las tres necesidades psicológicas jugaban un papel significativo en la predicción de la intención de ser físicamente activo. Los resultados de este estudio sugieren que es relevante fomentar la satisfacción de las NPB en los adolescentes argentinos para mejorar la adherencia a la práctica de actividad física.

En conclusión, los resultados de este estudio apoyan el uso de la BPNES en diversos contextos ofreciendo evidencia de la validación transcultural de la BPNES para utilizarse con estudiantes de EF de tres países que comparten el mismo idioma. Este hallazgo puede facilitar el desarrollo de futuros estudios que examinen las NPB junto a otras variables de actual relevancia. Este trabajo podría considerarse un estudio preliminar de las diferencias interculturales en la satisfacción de las NPB de estudiantes de EF de Argentina, España y Colombia, aunque se hacen necesarias más investigaciones en este ámbito con el fin de expandir el conocimiento sobre las diferencias transculturales en este campo. Por otro lado, los hallazgos del trabajo sugieren que los adolescentes argentinos presentan menor satisfacción con las NPB que los jóvenes españoles y colombianos.

Sería interesante que futuros estudios pro-

fundizaran en el análisis de la consistencia interna del factor de *competencia* en la BPNES para su uso en adolescentes latinoamericanos, modificando el ítem que ha mostrado un bajo peso de regresión en la muestra argentina y colombiana y volviendo a comprobar las propiedades psicométricas del instrumento. Por otra parte, parecen necesarias más investigaciones que ahonden en los motivos que expliquen los niveles inferiores en variables motivacionales que muestran los adolescentes argentinos con respecto a los jóvenes de otros países.

Entre las limitaciones del estudio, cabe mencionar la no aleatorización en la selección de la muestra y la ausencia de un análisis de validez concurrente a través del análisis de la asociación de las dimensiones de la BPNES con otros constructos teóricamente relacionados. Por último, es necesario señalar que en las diferencias entre países pudieron haber influido otros factores que no han sido considerados, tales como los contenidos que se trabajan o el estilo interpersonal del profesor.

La principal contribución de este estudio ha sido proporcionar evidencia de la validez y confiabilidad de la versión española de la BPNES para ser utilizada en Argentina y Colombia.

Referencias

- Aibar, A., Julián, J. A., Murillo, B., García-González, L., Estrada, S., & Bois, J. (2015). Actividad física y apoyo a la autonomía: El rol del profesor de educación física. *Revista de Psicología del Deporte*, 24(1), 155-161. Recuperado de http://www.rpd-online.com/article/view/aibar_solana_julian_etal
- Arbuckle, J. L. (2014). Amos (Versión 23.0) [software de cómputo]. Chicago: IBM SPSS.
- American Psychological Association (2002). *Ethical principles of psychologists and code of conduct (Amended June 2, 2010)*. American Psychological Association.
- Barclay, D., Higgins, C., & Thompson, R. (1995). The Partial Least Squares (PLS) approach to causal modelling: Personal computer adoption and use as an illustration. *Technology Studies*, 2(Special Issue on Research Methodology, 2), 285-309.
- Bentler, P. M. (1989). *EQS structural equations program manual*. Los Angeles, CA: BMDP Statistical Software.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.
- 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.
- Byrne, B. M. (2004). Testing for multigroup invariance using AMOS: A road less traveled. *Structural Equation Modeling*, 11(2), 272-300. doi: 10.1207/s15328007sem1102_8
- Cantú-Berrueto, A., Castillo, I., López-Walle, J., Tristán, J., & Balaguer, I. (2016). Estilo interpersonal del entrenador, necesidades psicológicas básicas y motivación: Un estudio en futbolistas universitarios mexicanos. *Revista Iberoamericana de Psicología del Ejercicio y el Deporte*, 11(2), 263-270. Recuperado de <http://www.webs.ulpgc.es/riped>
- Cid, L., Lettnin, C., Stobäus, C., Monteiro, D., Davoglio, T., & Moutão, J. (2016). Cross-cultural validation of the basic psychological needs in Physical Education Scale between Portugal and Brazil samples. *Spanish Journal of Psychology*, 19(e5), 1-10. doi: 10.1017/sjp.2016.6
- Chatzisarantis, N., & Hagger, M. (2009). Effects of an intervention based on self-determination theory on self-reported leisure-time physical activity participation. *Psychology and Health*, 24(1), 29-48. doi: 10.1080/08870440701809533
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, 14(3), 464-504. doi: 10.1080/10705510701301834

- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9(2), 233-255. doi: 10.1207/S15328007SEM0902_5
- Chin, W. W. (1998). Issues and opinion on structural equation modeling. *MIS Quarterly*, 22(1), 7-17.
- Cuevas, R., García-López, L. M., & Contreras, O. (2015). Influencia del modelo de Educación Deportiva en las necesidades psicológicas básicas. *Cuadernos de Psicología del Deporte*, 15(2), 155-162. Recuperado de <http://revistas.um.es/cpd/>
- Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behaviour*. New York: Plenum.
- Deci, E. L., & Ryan, R. M. (2000). The “What” and “Why” of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, 11(4), 227-268. doi: 10.1207/s15327965pli1104_01
- Deci, E. L., & Ryan, R. M. (2001). *Questionnaires: Basic Psychological Needs Scale*. Recuperado de <http://selfdeterminationtheory.org/basic-psychological-needs-scale/>
- Deci, E. L., Ryan, R. M., Gagné, M., Leone, D., Usunov, J., & Kornazheva, B. P. (2001). Need satisfaction, motivation, and well-being in the work organizations of a former eastern bloc country: A cross-cultural study of self-determination. *PSPB*, 27(8), 930-942. doi: 10.1177/0146167201278002
- Duda, J. L., & Allison, M. T. (1990). Cross-cultural analysis in exercise and sport psychology: A void in the field. *Journal of Sport and Exercise Psychology*, 12(2), 114-131. doi: 10.1123/jsep.12.2.114
- Duda, J. L., & Hayashi, C. (1998). Measurement issues in cross-cultural research within sport and exercise psychology. En J. L. Duda (Ed.), *Advances in sport and exercise psychology measurement* (pp. 471-483). Morgantown: Fitness Technology.
- Evelein, F., Korthagen, F., & Brekelmans, M. (2008). Fullfillment of the basic psychological needs of student teachers during their first teaching experiences. *Teaching and Teacher Education*, 24(5), 1137-1148. doi: 10.1016/j.tate.2007.09.001
- Fornell, C., & Larcker, D. F. (1981). Evaluating structural equations models with unobservable variables and measurement error. *Journal of Marketing Research*, 18, 39-50. doi: 10.2307/3151312
- Franco, E., Coterón, J., & Gómez, V. (2017). El rol de la motivación y la autoestima en la práctica de actividad física en adolescentes. *Revista Latinoamericana de Ciencia Psicológica*, 9(2), 1-15.
- Franco, E., Coterón, J., Gómez, V., Brito, J., & Martínez-González, A. (2017). Influencia de la motivación y del flow disposicional sobre la intención de realizar actividad físico-deportiva en adolescentes de cuatro países. *Retos*, 31, 46-51. Recuperado de <https://recyt.fecyt.es/index.php/retos>
- Franco, E., Coterón, J., Gómez, V., & De Franza, A. L. (2017). Relación entre motivación, actividad física realizada en el tiempo libre y la intención futura de práctica de actividad física. Estudio comparativo entre adolescentes argentinos y españoles. *SPORT TK*, 6(1), 25-34. Recuperado de <http://revistas.um.es/sportk>
- Franco, E., Coterón, J., & Martínez, H. A. (en revisión). Invarianza cultural del cuestionario de la orientación a la tarea y al ego y diferencias en las orientaciones motivacionales entre estudiantes de educación física de España, Argentina, Colombia y Ecuador. *Revista de Psicología del Deporte*.
- Franco, E., Coterón, J., Martínez, H. A., & Brito, J. (2017). Perfiles motivacionales en estudiantes de educación física de tres países y su relación con la actividad física. *Suma Psicológica*, 24(1), 1-8. doi: 10.1016/j.sumpsi.2016.07.001
- García-González, L., Aibar, A., Sevil, J., Almolda-Tomás, F. J., & Julián-Clemente, J. A. (2015). Soporte de autonomía en educación física: Evidencias para mejorar el proceso de enseñanza. *Cultura, Ciencia y Deporte*, 10(29), 103-111. doi: 10.12800/ccd.v10i29.547
- Gómez-Rijo, A., Hernández-Moreno, J., Martínez-Herráez, I., & Gámez, S. (2014). Necesidades psicológicas básicas en educación física según el género y el curso

- del estudiante. *Revista de Investigación Educativa*, 32(1), 159-167. doi: 10.6018/rie.32.1.172311
- González-Cutre, D., Martínez-Galindo, C., Alonso-Villodre, N., Cervelló-Gimeno, E., Conte-Marín, L., & Moreno-Murcia, J. A. (2007). Las creencias implícitas de habilidad y los mediadores psicológicos como variables predictoras de la motivación autodeterminada en deportistas adolescentes. En J. Castellano & O. Usabiaga (Eds.), *Investigación en la Actividad Física y el Deporte II* (pp. 407-417). Vitoria: Universidad del País Vasco.
- Gregorich, S. E. (2006). Do self-report instruments allow meaningful comparisons across diverse population groups? Testing measurement invariance using the confirmatory analysis framework. *Medical Care*, 44(Suppl 3), S78-S94. doi: 10.1097/01.mlr.0000245454.12228.8f
- He, J., & Van de Vijver, F. (2012). Bias and equivalence in cross-cultural research. *Online Readings in Psychology and Culture*, 2(2). doi: 10.9707/2307-0919.1111
- Herrero, J. (2010). El análisis factorial confirmatorio en el estudio de la estructura y estabilidad de los instrumentos de evaluación: Un ejemplo con el Cuestionario de Autoestima CA-14. *Psychosocial Intervention*, 19(3), 289-300. doi: 10.5093/in2010v19n3a9
- Hu, L., & Bentler, P. M. (1995). Evaluating model fit. En R. Hoyle (Ed.), *Structural equation modeling: Issues, concepts, and applications*. Newbury Park, CA: Sage.
- IBM Corp. (2016). IBM SPSS Statistics for Windows, Version 24.0. [software de cómputo] Armonk, NY: IBM Corp.
- López-Walle, J., Balaguer, I., Castillo, I., & Tristán, J. (2011). Clima motivacional percibido, motivación autodeterminada y autoestima en jóvenes deportistas mexicanos. *Revista de Psicología del Deporte*, 20(1), 209-222. Recuperado de <http://www.redalyc.org/html/2351/235119302015/>
- López-Walle, J., Tomás, I., Castillo, I., Tristán, J., & Balaguer, I. (2011). Invarianza factorial del TEOSQ en jóvenes deportistas mexicanos y españoles. *Revista Mexicana de Psicología*, 28, 53-61. Recuperado de <http://eprints.uanl.mx/5312>
- Marsh, H. W., Hau, K. T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling*, 11(3), 320-341. doi: 10.1207/s15328007sem1103_2
- Méndez-Giménez, A., Fernández-Río, J., & Cecchini-Estrada, J. A. (2013). Papel importante del alumnao, necesidades psicológicas básicas, regulaciones motivacionales y autoconcepto físico en educación física. *Cuadernos de Psicología del Deporte*, 13(1), 71-82. doi: 10.4321/s1578-84232013000100008
- Meneses-Montero, M., & Ruiz-Juan, F. (2017). Estudio longitudinal de los comportamientos y el nivel de actividad físico-deportiva en el tiempo libre en estudiantes de Costa Rica, México y España. *Retos*, 31, 219-226. Recuperado de <https://dialnet.unirioja.es/servlet/articulo?codigo=5841375>
- Millsap, R. E., & Kwok, O. (2004). Evaluating the impact of partial factorial invariance on selection in two populations. *Psychological Methods*, 9(1), 93-115. doi: 10.1037/1082-989X.9.1.93
- Moreno-Murcia, J. A., Martínez-Galindo, C., Moreno-Pérez, V., Marcos, P. J., & Borges, F. (2012). Confirmation of the Basic Psychological Needs in Exercise Scale (BPNES) with a sample of people who do healthy exercise. *Journal of Sports Science and Medicine*, 11, 141-146. Recuperado de <http://www.jssm.org>
- Moreno-Murcia, J. A., González-Cutre-Coll, D., Chillón-Garzón, M., & Parra-Rojas, N. (2008). Adaptación a la educación física de la Escala de las Necesidades Psicológicas Básicas en el Ejercicio. *Revista Mexicana de Psicología*, 25(2), 295-303. Recuperado de <http://www.redalyc.org/pdf/2430/243016308009.pdf>
- Moutão, J., Serra-Majem, L., Alves, J. A., Leitao, J. C., & Vlachopoulos, S. (2012). Validation of the Basic Psychological Needs in Exercise Scale in a

- Portuguese sample. *Spanish Journal of Psychology*, 15(1), 399-409. doi: 10.5209/rev_sjop.2012.v15.n1.37346
- Nicholls, J. G. (1984a). Achievement motivation: Conceptions of ability, subjective experience, task choice, and performance. *Psychological Review*, 91(3), 328-346. doi: 10.1037/0033-295x.91.3.328
- Nicholls, J. G. (1984b). Conceptions of ability and achievement motivation. En R. E. Ames & C. Ames (Eds.), *Research on motivation in education, Vol. 1: Student motivation* (pp. 39-73). New York: Academic Press.
- Ntoumanis, N. (2005). A prospective study of participation in optional school physical education using a Self-Determination Theory framework. *Journal of Educational Psychology*, 97(3), 444-453. doi: 10.1037/0022-0663.97.3.444
- Pulido, J. J., Sánchez-Oliva, D., González-Ponce, I., Amado, D., & Chamorro, J. L. (2017). Adaptación al fútbol de la Escala de Necesidades Psicológicas Básicas en el Ejercicio. *Revista Iberoamericana de Psicología del Ejercicio y el Deporte*, 11(2), 177-184. Recuperado de <http://www.redalyc.org/articulo.oa?id=311145841001>
- Raykov, T., & Marcoulides, G. A. (2006). *A first course in structural equation modeling*. Mahwah, NJ: Erlbaum.
- Reyes, M., Raimundi, M. J., & Gómez-Correa, L. (2012). Programa de entrenamiento en habilidades psicológicas en jugadoras de voleibol de alto rendimiento. *Cuadernos de Psicología del Deporte*, 12(1), 9-16. doi: 10.4321/s1578-84232012000100001
- Ryan, R. M. (1991). The nature of the self in autonomy and relatedness. En J. Strauss & G. R. Goethals (Eds.), *Multidisciplinary perspectives on the self* (pp. 208-238). New York: Springer-Verlag.
- Sánchez, J., & Núñez, J. L. (2007). Análisis preliminar de las propiedades psicométricas de la versión española de la Escala de Necesidades Psicológicas Básicas en el Ejercicio Físico. *Revista Iberoamericana de Psicología del Ejercicio y el Deporte*, 2(2), 83-92. Recuperado de <http://www.redalyc.org/articulo.oa?id=311126258006>
- Sevil, J., Abós, A., Julián-Clemente, J. A., Murillo, B., & García-González, L. (2015). Género y motivación situacional en educación física: Claves para el desarrollo de estrategias de intervención. *Revista Internacional de Ciencias del Deporte*, 11(41), 281-296. doi: 10.5232/ricyde2015.04106
- Standage, M., Sebire, S. J., & Loney, T. (2008). Does exercise motivation predict engagement in objectively assessed bouts of moderate-intensity exercise?: A Self-Determination Theory perspective. *Journal of Sport and Exercise Psychology*, 30(4), 337-352. doi: 10.1123/jsep.30.4.337
- Van Aart, I., Hartman, E., Elferink-Gemser, M., Mombarg, R., & Visscher, C. (2017). Relations among basic psychological needs, PE-motivation and fundamental movement skills in 9-12-year-old boys and girls in Physical Education. *Physical Education and Sport Pedagogy*, 22(1), 15-20. doi: 10.1080/17408989.2015.1112776
- Verloigne, M., Lippevelde, W. V., Maes, L., Yildirim, M., Chinapaw, M., Manios, Y. ... De Bourdeaudhuij, I. (2012). Levels of physical activity and sedentary time among 10- to 12- year-old boys and girls across 5 European countries using accelerometers: An observational study within the ENERGY-project. *International Journal of Behavioral Nutrition and Physical Activity*, 9(1), 34. doi: 10.1186/1479-5868-9-34
- Vlachopoulos, S. (2008). The Basic Psychological Needs in Exercise Scale: Measurement invariance over gender. *Structural Equation Modeling*, 15(1), 114-135. doi: 10.1080/10705510701758398
- Vlachopoulos, S., Asci, F. H., Cid, L., Ersoz, G., González-Cutre, D., Moreno-Murcia, J. A., & Moutao, J. (2013). Cross-cultural invariance of the Basic Psychological Needs in Exercise Scale and need satisfaction latent mean differences among Greek, Spanish, Portuguese and Turkish samples. *Psychology of Sport & Exercise*, 14(5), 622-631. doi: 10.1016/j.psychsport.2013.03.002

- Vlachopoulos, S., & Michailidou, S. (2006). Development and initial validation of a measure of autonomy, competence and relatedness: the Basic Psychological Needs in Exercise Scale. *Measurement in Physical Education & Exercise Science*, 10(3), 179-201. doi: 10.1207/s15327841mpee1003_4
- Vlachopoulos, S., Ntoumanis, N., & Smith, A. (2010). The Basic Psychological Needs in Exercise Scale: Translation and evidence for cross-cultural validity. *International Journal of Sport and Exercise Psychology*, 8(4), 394-412. doi: 10.1080/1612197x.2010.9671960
- Wang, J. C. K., Pyun, D. Y., Kim, J. Y., & Chatzisarantis, N. (2009). Testing for multigroup invariance of the perceived locus of causality in sport. *Personality and Individual Differences*, 47(6), 590-597. doi: 10.1016/j.paid.2009.05.008
- Wilson, P., Rogers, W., Rodgers, W., & Wild, T. (2006). The Psychological Need Satisfaction in Exercise Scale. *Journal of Sport and Exercise Psychology*, 28(3), 231-251. doi: 10.1123/jsep.28.3.231
- Zamarripa, J., Castillo, I., Tomás, I., & López-Walle, J. (2016). Validación mexicana del Cuestionario de Apoyo a las Necesidades Psicológicas Básicas en la Educación Física. *Revista Mexicana de Psicología*, 33(2), 143-150.
-

Validación de la Escala de Gravedad de Fatiga en Población General de la Ciudad de México

Validation of the Fatigue Severity Scale in general population of Mexico City

Eduardo Velasco-Rojano *¹, Rocío E. Duarte-Ayala², Angelica Riveros-Rosas³,
Juan José Sánchez-Sosa¹, L. Isabel Reyes-Lagunes¹

1 - Universidad Nacional Autónoma de México, Facultad de Psicología, México.

2 - Universidad del Valle de México, Campus Lomas Verdes, México.

3 - Universidad Nacional Autónoma de México, Facultad de Contaduría y Administración, México.

Recibido: 07/08/2017 Revisado: 24/08/2017 Aceptado: 04/09/2017

Introducción
Objetivos
Justificación
Método
Resultados
Discusión
Conclusión
Referencias

Resumen

Para conocer la prevalencia y las consecuencias de la fatiga se necesitan formas confiables y válidas para medirla. Una de las escalas más usadas es la Escala de Gravedad de Fatiga (FSS; Krupp, LaRocca, Muir-Nash, & Steinberg, 1989), por lo que el propósito de este trabajo fue validar esta escala psicométricamente de manera culturalmente relevante, identificar si su estructura es la más adecuada a través de un análisis factorial confirmatorio y comprobar la invarianza factorial por sexo. Se trabajó con dos grupos de participantes, 150 habitantes de la Ciudad de México para la validación y 220 habitantes de la Ciudad de México para el análisis factorial confirmatorio y prueba de invarianza. Se encontró una estructura con seis reactivos, adecuada confiabilidad ($\alpha = .93$) correcto ajuste [$\chi^2_{(9)} = 12.56, p = .183$; CFI = .99; RMSEA = .05] e invarianza por sexo, por lo que se concluye que la Escala de Gravedad de Fatiga es una medida confiable y válida.

Palabras clave: *fatiga, medición, validación psicométrica, invarianza, sexo*

Abstract

Reliable and valid ways to measure fatigue are required in order to determine its prevalence and consequences. One of the methods widely used to measure fatigue is the Fatigue Severity Scale (FSS; Krupp, LaRocca, Muir-Nash, & Steinberg, 1989). Thus, the purpose of this study was to validate this scale psychometrically in a culturally relevant manner, to identify whether its structure is the most appropriate through a confirmatory factor analysis, and to test the invariance by sex. Two groups of participants, inhabitants of Mexico City, were included in the present study, 150 for validation studies, and 220 for confirmatory factor analysis and invariance test studies. A structure with six items, adequate reliability ($\alpha = .93$), goodness of fit [$\chi^2_{(9)} = 12.56, p = .183$; CFI = .99; RMSEA = .05] and invariance by sex was found. So, it is concluded that FSS is a reliable and valid measure of fatigue.

Key words: *fatigue, assessment, psychometric validation, invariance, gender*

*Correspondencia a: Eduardo Velasco Rojano. Universidad Nacional Autónoma de México, Facultad de Psicología, Edificio D, Cubículo 10 Mezzanine. Avenida Universidad 3004, Coyoacán, Copilco Universidad, 04510 Ciudad de México, CDMX. México. Teléfono: 0155 56 22 23 23. eduardorojanova@gmail.com

Cómo citar este artículo: Velasco-Rojano, E., Duarte-Ayala, R. E., Riveros-Rosas, A., Sánchez-Sosa, J. J., & Reyes-Lagunes, L. I. (2017). Validación de la Escala de Gravedad de Fatiga en población general de la Ciudad de México. *Revista Evaluar*, 17(2), 113-125. Recuperado de <https://revistas.unc.edu.ar/index.php/revaluar>

Nota del autor: Una versión preliminar de la adaptación aplicada a otra muestra se presentó en el XVI Congreso Mexicano de Psicología Social. Para este estudio, la escala se estudió en participantes distintos en dos nuevas muestras, por lo que los aquí presentados son datos diferentes y no hay conflicto de interés.

Introducción

La *fatiga* puede entenderse como la sensación de cansancio físico y falta de energía que ocurre como consecuencia de la realización de esfuerzos físicos o mentales o de la existencia de enfermedad (Gottschalk et al., 2005; Viner & Christie, 2005). La fatiga prolongada puede tener consecuencias en las relaciones sociales; en la salud mental, al asociarse con estrés y depresión; en la memoria, y en la motricidad fina (Malta-Purim, Bittencourt-Gumarães, Kapp-Titski, & Leite, 2016).

Existe confusión alrededor del término fatiga ya que se presenta ligado con otros como astenia, cansancio, agotamiento, apatía, aburrimiento, psicastenia, pesimismo y vulnerabilidad, como señalan Fuentes-Márquez, Senín-Calderón, Rodríguez-Testal y Carrasco (2015). Esta situación se agrava al considerar que la definición del concepto es imprecisa y no existe un consenso general acerca de la posibilidad de considerarse un proceso adaptativo de las personas sanas, un síntoma físico, un síntoma psicológico o un síndrome (Fuentes-Márquez et al., 2015).

Al considerar esta información, puede entenderse que no sea fácil caracterizar apropiadamente el alcance de las molestias y consecuencias de la fatiga, ni su prevalencia en la población general, lo que se acentúa porque la mayoría de las quejas por fatiga suelen ser desestimadas ya sea por los propios pacientes o por los médicos que las escuchan. Esto se debe a que se trata de un padecimiento con síntomas no específicos, con lo que se considera que alrededor del 70 por ciento de los casos en etapas tempranas permanecen sin detectar y atender (Van't Leven, Zielhuis, Van der Meer, Verbeek, & Bleijenberg, 2009). Así, es necesario realizar estudios para conocer la prevalencia y las consecuencias de la fatiga en la población general y, además, para realizar este tipo de investigación es necesario contar con formas

confiables y válidas para medirla.

La fatiga puede ser evaluada de dos maneras distintas: con aparatos para realizar distintas pruebas físicas (Ali et al., 1994) o con escalas de autoinforme (Gottschalk et al., 2005; Kos et al., 2005), que es la que más se utiliza dado que las mediciones con aparatos son más complejas y costosas (Gottschalk et al., 2005; Kos et al., 2005).

Sin embargo, el desarrollo de escalas de autoinforme es difícil debido a la variabilidad en la experiencia de las molestias informadas (Armutlu et al., 2007; Gottschalk et al., 2005; Kos et al., 2005). Existe un amplio número de instrumentos diseñados para medir la fatiga, pero no existe suficiente investigación sobre sus propiedades psicométricas (Whitehead, 2009). En el caso particular de la población hispanohablante, la falta de estudios sobre los instrumentos hasta ahora adaptados y/o utilizados en experiencias empíricas de diagnóstico y medición también ha llevado a un vacío en la investigación sobre la relación de los puntajes de fatiga con otras variables y sus posibles consecuencias.

En la población hispanohablante se han utilizado diversos instrumentos para medir la fatiga, como el Test de Patrones Subjetivos de Fatiga (PSF; Yoshitake, 1971), el cual se ha empleado para evaluar fatiga en ambientes laborales en México (Orrantia-Daniel & De la Vega-Bustillos, 2003) y Colombia (Neisa, Tangarife, & Godoy, 2014). El PSF consta de 30 reactivos dicotómicos que se responden por *sí* o por *no*, y miden la percepción subjetiva de la fatiga laboral en tres dimensiones: la exigencia mental, los síntomas físicos y los síntomas mixtos.

La subescala de fatiga subjetiva de la Checklist for Individual Strength (CIS; Beurskens et al., 2000), usada en Colombia para evaluar fatiga en ambientes laborales (Useche, Cendales, & Gómez, 2017), mide percepción de fatiga subjetiva relacionada con aspectos conductuales y está

conformada por ocho reactivos con siete opciones de respuesta sobre grado de acuerdo con la situación actual del participante. El coeficiente alfa de Cronbach de la escala fue de $\alpha = .86$. La Need For Recovery Scale (NFR; De Croon, Sluiter, & Frings-Dresen, 2006) también ha sido utilizada para evaluar fatiga en ambientes laborales en Colombia (Useche et al., 2017) y mide los síntomas de fatiga física y mental con 11 reactivos dicotómicos con una confiabilidad de $\alpha = .80$.

La Chronic Fatigue Scale (CFS; Chalder et al., 1993), que mide las dos dimensiones de fatiga física y fatiga mental, con 11 reactivos tipo Likert con cuatro opciones de respuesta y una consistencia interna (α) entre .88 y .90, se ha utilizado para comparar el síndrome de fatiga crónica en población anglosajona e hispanohablante en Estados Unidos (Torres-Harding, Mason-Schutter, & Jason, 2008).

Otro instrumento de los más utilizados es la Escala de Gravedad de la Fatiga (FSS; Krupp et al., 1989), la cual mide gravedad de fatiga de manera unidimensional con nueve reactivos tipo Likert con siete opciones de respuesta en términos de acuerdo y desacuerdo. Esta escala se ha utilizado y validado psicométricamente en población general y en pacientes con distintos tipos de enfermedades crónicas, como esclerosis múltiple, lupus, hepatitis C y pacientes con VIH (Lerdal et al., 2011). Además, la escala FSS se ha adaptado a diversos idiomas, como el turco (Gök, Cengiz, Erol, & Özgöçmen, 2016), el italiano (Ottonello, Pellicciari, Giordano, & Foti, 2016) y el español colombiano (Bernal-Vargas, Riveros-Munévar, Vinaccia-Alpi, & Quiceno-Sierra, 2017), y en todos los casos se encontraron adecuadas propiedades psicométricas en términos de discriminación, confiabilidad y validez (Whitehead, 2009).

La ventaja principal de esta escala es su breve extensión. Los nueve reactivos originales miden si la fatiga afecta la motivación, el ejer-

cicio y el funcionamiento físico, las actividades de la vida cotidiana y social. Existen versiones con siete reactivos desarrolladas en pacientes con VIH y esclerosis múltiple angloparlantes que no incluyen las consecuencias de la fatiga en la motivación o el ejercicio físico (Johansson, Kottorp, Lee, Gay, & Lerdal, 2014; Lerdal et al., 2011). Por otra parte, existe una versión con ocho reactivos evaluada en pacientes de esclerosis múltiple italo-parlantes que no incluye consecuencias sobre la motivación (Ottonello et al., 2016).

La FSS ha sido utilizada previamente en México para evaluar fatiga en programas de educación médica (Duarte-Ayala, 2017), pero no existen estudios de validación en población general de la Ciudad de México.

Objetivos

El presente estudio tuvo como objetivos: validar psicométricamente una versión adaptada culturalmente relevante de la Escala de Gravedad de Fatiga en población general de la Ciudad de México, comprobar cuál es la estructura factorial más adecuada con análisis factorial confirmatorio, y realizar una prueba de invarianza por sexo para determinar si su funcionamiento difiere en hombres y mujeres.

Justificación

Con el presente estudio se validará el instrumento de fatiga FSS a fin de determinar su prevalencia en la población general de la Ciudad de México. Además, nuestros resultados pueden ser válidos para el diagnóstico de la fatiga en etapas tempranas, lo que hasta el momento es una necesidad identificada por la literatura del campo (Van't Leven et al., 2009).

Método

Tipo de estudio

Se llevó a cabo un estudio instrumental psicométrico (Montero & León, 2007).

Participantes

Se trabajó con dos grupos de participantes. El grupo de validación psicométrica estuvo conformado por 180 habitantes de la Ciudad de México, 95 mujeres y 85 hombres, entre 18 y 60 años, con una edad promedio de 29.84 años y una desviación estándar de 10.75. Con respecto a la escolaridad de los participantes, 10 habían finalizado la escuela primaria, 14 la escuela secundaria, 35 la preparatoria, 100 la licenciatura y 21 habían culminado estudios de posgrado. En cuanto al estado civil, 132 eran solteros y 48 eran casados. Para determinar el tamaño de la muestra se decidió exceder 10 veces el número de reactivos, de acuerdo con la propuesta de Nunnally y Bernstein (1994), lo que se corroboró con un adecuado valor ($> .60$) de la medida de adecuación muestral Kaiser-Meyer-Olkin.

El grupo de análisis factorial confirmatorio y prueba de invarianza estuvo conformado por 220 habitantes de la Ciudad de México, 110 mujeres y 110 hombres, entre 18 y 60 años, con un promedio de 32.33 años y una desviación estándar de 11.60. Respecto a la escolaridad de los participantes, 7 habían finalizado la primaria, 22 la secundaria, 65 la preparatoria, 107 la licenciatura y 19 un posgrado, 110 eran solteros y 110 casados. El tamaño de la muestra se consideró con base en un análisis de potencia estadística calculado para GFI (MacCallum & Hong, 1997), en el que se consideró tener nueve reactivos, 45 grados de libertad, un coeficiente alfa de Cronbach de .05 y una potencia deseada de .80, lo que resultó en

157 participantes. Se decidió exceder 20 veces el número de reactivos, de acuerdo con la propuesta de Jackson (2003).

Instrumento

Para la validación psicométrica se utilizó la versión traducida y adaptada de la Escala de Gravedad de Fatiga (FSS; Krupp et al., 1989), producto de un estudio previo (Duarte, Velasco, & Reyes-Lagunes, 2016), la cual se compone de nueve afirmaciones con siete opciones de respuesta de tipo Likert pictórico (Reyes-Lagunes, 1993), rotuladas *De acuerdo* y *En desacuerdo* en los extremos.

Para el análisis factorial confirmatorio y prueba de invarianza se empleó la versión validada de la Escala de Gravedad de Fatiga (FSS; Krupp et al., 1989) producto de la fase anterior, con nueve reactivos y un coeficiente alfa de Cronbach de .93.

También se utilizó la Escala de Impacto de Fatiga (FIA; Duarte-Ayala, Velasco-Rojano, Sánchez-Sosa, & Reyes-Lagunes, 2017), validada en población mexicana con un coeficiente alfa de Cronbach de .93, con el fin de comprobar validez convergente.

Procedimiento

En la validación psicométrica, los participantes fueron convocados a lugares públicos, como jardines, parques, plazas y mercados; se obtuvo su consentimiento, se les administró el instrumento, leyendo cada reactivo y las opciones de respuesta en los casos en que esto fuera necesario, como con participantes invidentes o que no supieran leer y escribir, y se les agradeció su participación. Una vez obtenidos los datos, se

Tabla 1

Análisis de Distribución de los Reactivos de la Escala de Gravedad de Fatiga FSS adaptada.

Reactivo	Análisis de distribución de frecuencia de los Reactivos							Prueba de Normalidad			
	Frecuencia en porcentaje por opción de respuesta							Asimetría	Curtosis	Z	p con corrección de Lilliefors
	En Desacuerdo			De acuerdo							
FSS1	14.44	11.67	13.33	14.44	17.22	17.22	11.67	-.14	-1.19	.14	$p < .01^{**}$
FSS2	20	21.67	13.89	17.78	15.56	7.22	3.89	.35	-.92	.18	$p < .01^{**}$
FSS3	21.11	17.22	15.56	20.56	10.56	8.89	6.11	.36	-.87	.15	$p < .01^{**}$
FSS4	15.56	17.22	21.67	16.11	13.33	8.33	7.78	.35	-.83	.15	$p < .01^{**}$
FSS5	21.67	20.56	16.67	17.78	12.78	5.56	5	.47	-.69	.17	$p < .01^{**}$
FSS6	19.44	19.44	17.22	16.67	13.89	8.33	5	.36	-.88	.16	$p < .01^{**}$
FSS7	21.67	20.56	16.11	12.78	11.11	11.67	6.11	.42	-1.03	.18	$p < .01^{**}$
FSS8	21.67	20.56	17.78	12.78	11.11	7.78	8.33	.51	-.86	.17	$p < .01^{**}$
FSS9	24.44	20.56	13.33	9.44	11.67	10.56	10	.44	-1.16	.20	$p < .01^{**}$

Nota. $**p \leq .01$

realizó la validación psicométrica culturalmente relevante propuesta por Reyes-Lagunes y García y Barragán (2008), que incluye análisis de distribución, discriminación, direccionalidad, confiabilidad y estructura factorial de los reactivos de la escala.

Para el análisis factorial confirmatorio y prueba de invarianza se convocó a los participantes como ha sido descrito anteriormente. Una vez obtenidos los datos se analizó cuál era la estructura factorial más adecuada y se realizó una prueba de invarianza por sexo siguiendo el procedimiento propuesto por Milsap y Olivera-Aguilar (2012), que consiste en evaluar cuatro modelos anidados, cada uno más restrictivo que el anterior. Primeramente, se probó el ajuste de la configuración teniendo en cuenta que la estructura factorial fuera la misma para los grupos. Posteriormente, se consideró la invarianza métrica o débil (en la que las cargas factoriales se restringen a igualdad), luego se evaluó el modelo con invarianza fuerte (en el que las cargas factoriales y los interceptos se restringen a igualdad), y finalmente un

modelo de invarianza estricta (en el que las cargas factoriales, los interceptos y las varianzas únicas de los reactivos se restringieron a igualdad) y se comparó el ajuste de los modelos. Si no se obtiene una diferencia significativa en chi cuadrada ni cambios mayores de .01 en el CFI, se considera que existe invarianza (Kline, 2016). Si se obtiene una invarianza estricta, los cambios resultan de la variable latente y no de un sesgo en la medición (DeShon, 2004).

Resultados

Validación Psicométrica

Se realizó un análisis de frecuencias de cada uno de los reactivos y se encontró que todas las opciones de respuesta tuvieron frecuencia. Posteriormente, se calcularon el sesgo, la curtosis y la prueba Z de Kolmogorov-Smirnov. Los resultados muestran que todos tuvieron una distribución sesgada, lo que fue considerado en análisis subsiguientes. La Tabla 1 incluye los porcentajes

Tabla 2
Análisis de Discriminación de los Reactivos de FSS versión adaptada.

Reactivo	Comparación de grupos extremos			Correlación reactivo total	Correlación reactivo total sin el reactivo
	U Mann Whitney	<i>gl</i>	<i>p</i>		
FSS1	85.00	45	$p < .01^{**}$.68**	.59**
FSS2	169.00	45	$p < .01^{**}$.70**	.62**
FSS3	12.00	45	$p < .01^{**}$.83**	.78**
FSS4	10.00	45	$p < .01^{**}$.84**	.79**
FSS5	17.50	45	$p < .01^{**}$.85**	.80**
FSS6	35.50	45	$p < .01^{**}$.84**	.80**
FSS7	0	45	$p < .01^{**}$.87**	.83**
FSS8	10.50	45	$p < .01^{**}$.84**	.79**
FSS9	1.00	45	$p < .01^{**}$.85**	.79**

Nota. ** $p \leq .01$.

por opción de respuesta, el coeficiente de asimetría, la curtosis y los resultados de la prueba de normalidad.

Posteriormente se estudió el poder discriminativo de los reactivos con dos análisis: una comparación de grupos extremos respecto al total de la prueba (puntajes inferiores al primer cuartil y puntajes superiores al tercer cuartil), mediante la prueba U de Mann Whitney, y una correlación entre el total de la prueba y cada uno de los reactivos. También se obtuvo/analizó una versión corregida de la misma en la cual el total de la prueba no incluía el reactivo con el que se iba a relacionar. Se eliminaron los reactivos con correlaciones bajas ($r < .30$) y todos los reactivos discriminaron significativamente como se muestra en la Tabla 2 (Magnusson, 1998).

Al analizar las correlaciones reactivo-total se verificó que en caso de eliminar cualquiera de los reactivos la consistencia de la escala disminuiría, debido a lo cual se consideró conveniente conservar la totalidad de los reactivos para mantener una buena consistencia. En la Tabla 2 se muestra

el valor de la comparación entre grupos extremos y las correlaciones entre reactivo y total.

A fin de conocer la estructura de la escala se realizó un análisis factorial exploratorio de ejes principales con rotación oblicua, adecuado para identificar las variables latentes que subyacen a los reactivos (Fabrigar, MacCallum, Wegener, & Strahan, 1999; Schmitt, 2011). Se eligió este método de extracción debido a que no se basa en el supuesto de normalidad, y se eligió el tipo de rotación debido a que este es más preciso y provee más información que la rotación ortogonal. Se encontró una solución unidimensional con los nueve reactivos que explica el 62% de la varianza.

En la Tabla 3 se muestran las cargas factoriales de cada uno de los reactivos, el autovalor y el porcentaje de varianza explicada.

Para verificar la pertinencia del análisis, se calculó la medida de adecuación muestral Kaiser-Meyer-Olkin ($KMO = .91$) y la prueba de esfericidad de Bartlett ($\chi^2_{(36)} = 1323.33, p < .01$) y se pudo concluir que la matriz de correlaciones de la que partía era adecuada. Posteriormente, se

calculó la confiabilidad total de la escala alfa de Cronbach y se obtuvo un valor adecuado de consistencia interna ($\alpha = .93$).

Tabla 3

Resultados del análisis factorial exploratorio de ejes principales con rotación oblicua del FSS adaptado.

Reactivo	Factor 1. Fatiga
FSS7	.87
FSS5	.84
FSS6	.84
FSS8	.83
FSS9	.83
FSS4	.82
FSS3	.81
FSS2	.64
FSS1	.60
% Varianza Explicada	.62
Autovalor o valor Eigen	5.66

Análisis factorial confirmatorio y prueba de invarianza

A continuación se llevó a cabo un análisis factorial confirmatorio de máxima verosimilitud robusta, método seleccionado por la falta de normalidad en los datos (West, Taylor, & Wu, 2012)

y se probaron tres modelos comparativos, todos con una dimensión pero con diferente número de reactivos: uno con los nueve reactivos, otro con ocho y uno con siete reactivos. La identificación de los modelos se basó en establecer la carga factorial del primer reactivo como 1 para definir la métrica de la variable latente (Kenny & Milan, 2012) con el software R y el paquete Lavaan (Rosseel, 2012).

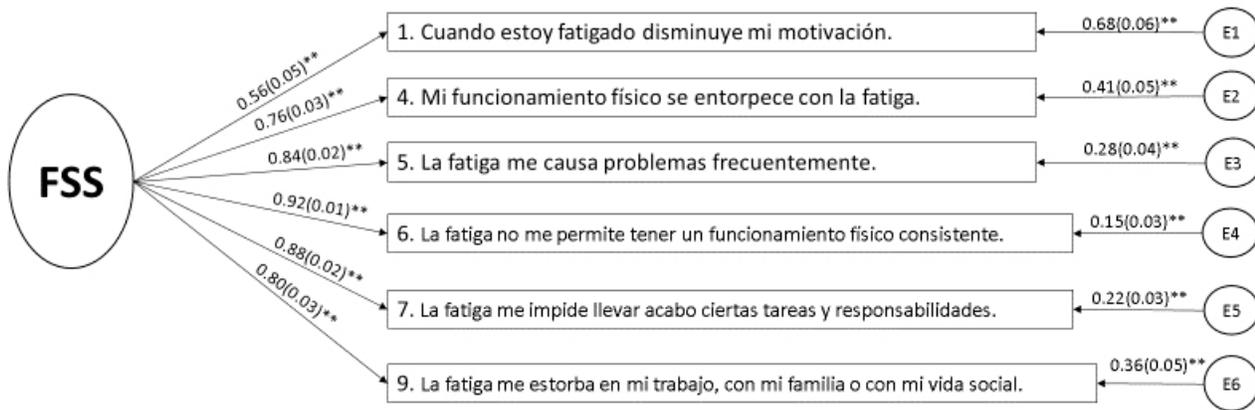
Para evaluar el ajuste de los modelos se utilizaron los siguientes índices de bondad de ajuste: Chi cuadrada normada, dividida por los grados de libertad (debería ser menor a 3.00 con una probabilidad no significativa; entre menor cociente mejor ajuste). Índice de bondad de ajuste Lisrel (GFI), índice de ajuste comparativo de Bentler (CFI; para que exista un buen ajuste los valores de GFI y CFI deben ser $\geq .95$), raíz cuadrada media del error de aproximación (RMSEA) y SRMR (los valores de RMSEA y SRMR deben ser $\leq .08$ para indicar un buen ajuste).

Los resultados de los tres modelos mostraron falta de ajuste global para explicar los datos, por lo que se analizaron las faltas de ajuste local mediante el uso de índices de modificación y se encontraron asociados con los reactivos 2, 3 y 8. A partir de estos resultados, se decidió modificar el modelo eliminando estos reactivos, ya que alcanzar ajuste mediante la inclusión de los mismos puede reflejar más la idiosincrasia de la muestra que la relación entre variables (MacCallum, Roznowski, & Necowitz, 1992). El modelo modi-

Tabla 4

Índices de ajuste para el análisis factorial confirmatorio de los modelos.

Modelo	χ^2	<i>p</i>	χ^2/gl	CFI	GFI	RMSEA	IC 90%	SRMR
9 reactivos	$\chi^2_{(27)} = 85.39$	$p < .01^{**}$	3.16	.94	.92	.12	[.09, .15]	.04
8 reactivos	$\chi^2_{(20)} = 64.01$	$p < .01^{**}$	3.20	.95	.93	.11	[.08, .14]	.04
7 reactivos	$\chi^2_{(14)} = 34.39$	$p < .01^{**}$	2.45	.97	.96	.09	[.05, .13]	.03
6 reactivos	$\chi^2_{(9)} = 12.56$	$p = .18$	1.39	.99	.99	.05	[0, .11]	.02

**Figura 1**

Análisis factorial confirmatorio de la FSS.

Nota. Solución ajustada

ficado con una dimensión y seis reactivos mostró un adecuado ajuste en todos los índices, como se muestra en la Tabla 4, que incluye los índices de ajuste para todos los modelos.

Las cargas factoriales estandarizadas de la escala, que resultan significativas, positivas y con valores adecuados para la solución ajustada, se muestran en la Figura 1.

Para determinar la existencia de un funcionamiento diferencial de los reactivos de acuerdo con el parámetro sexo del individuo, se realizó una prueba de invarianza factorial de medición. Los resultados muestran una invarianza factorial estricta, ya que no se encontraron diferencias sig-

nificativas en el ajuste de los modelos anidados, tal como lo indica la Tabla 5.

La Tabla 5 incluye los datos de la prueba de diferencias de la χ^2 y cambio en CFI para los modelos anidados.

A continuación, se compararon ambos grupos en la media de la variable latente *fatiga* y no se encontraron diferencias estadísticamente significativas ($Z = -0.35, p = .72$).

Como evidencia de validez concurrente, se calculó la correlación con el puntaje total de la Escala de Impacto de Fatiga y se encontró una correlación media estadísticamente significativa ($r = .59, p < .001$).

Tabla 5

Prueba de diferencia de la χ^2 en los modelos de la dimensión *fatiga*.

Modelo	χ^2	<i>gl</i>	$\Delta\chi^2$	Δgl	<i>p</i>	CFI	ΔCFI
Configuración	25.36	18	-	-	-	0.99	-
Métrica	27.20	23	2.03	5	.84	1	.01
Fuerte	32.21	28	4.89	5	.42	1	0
Estricta	39.37	34	5.10	6	.53	1	0

Discusión

Uno de los objetivos del presente trabajo era validar psicométricamente de manera culturalmente relevante la Escala de Gravedad de la Fatiga en población general de la Ciudad de México. Para ello se buscó partir de una traducción adecuada al lenguaje que no sacrificara la evaluación de los constructos que la escala original propone. La versión final después de su aplicación y análisis cuenta con seis reactivos, un nivel de confiabilidad de $\alpha = .91$ y todos los reactivos se encuentran relacionados significativamente, lo que corrobora que se trata de la medición de un mismo constructo.

Mediante la validación psicométrica a través del análisis factorial exploratorio de ejes principales, se encontró que la estructura que emerge libremente de la población tiene una dimensión con nueve reactivos como en la estructura de la escala original (Krupp et al., 1989) con una adecuada consistencia interna. Sin embargo, en el análisis factorial confirmatorio se encontró que los reactivos 2 (*El ejercicio aumenta mi fatiga*), 3 (*Me fatigo fácilmente*) y 8 (*La fatiga es uno de los tres síntomas que menos me permiten hacer cosas*) no eran adecuados.

La falta de ajuste de los reactivos 2 y 3 que indican problemas en el funcionamiento físico puede explicarse por dos posibles razones, la primera es que son reactivos que surgieron en poblaciones afectadas por enfermedades crónicas, en particular en pacientes con esclerosis múltiple que entre sus principales síntomas tienen debilidad muscular en las extremidades y dificultad con la coordinación y el equilibrio (Instituto Nacional de Desórdenes Neurológicos y Accidentes Vasculares [NINDS], 2002), lo que no corresponde con el estado de salud de la población general, la segunda es que contestar afirmativamente a estos reactivos puede implicar una disminución de

la capacidad de trabajo, lo que en México no es socialmente aceptado por ser una cultura que valora mucho el trabajo (Díaz-Guerrero, 2003), esta razón también puede explicar que el reactivo 8 no funcione según se esperaba dado que se refiere a la capacidad de hacer cosas.

La extensión de la escala la convierte en una versión corta que incluye sólo los reactivos con las mejores propiedades psicométricas para la población meta. Haber obtenido una versión corta es una ventaja, ya que es más práctica en su aplicación. Es importante destacar que en el presente trabajo se obtuvo una nueva versión de la FSS, por lo que los resultados obtenidos no son comparables con los resultados obtenidos con otras versiones de la misma escala, ya que al cambiar el número de reactivos, la definición del constructo cambia y se circunscribe a los resultados obtenidos en población general de la Ciudad de México.

Una evidencia de validez concurrente es la correlación media con la Escala de Impacto de Fatiga, la cual es importante porque se trata de una escala que mide fatiga en la misma población meta con criterios rigurosos de confiabilidad y validez, además de que ha sido utilizada en otras poblaciones para probar este tipo de validez (Oncu, Atamaz, Durmaz, & On, 2013).

En cuanto al objetivo de comparar por sexo, se observó que existe invarianza factorial estricta entre hombres y mujeres, lo que significa que no existe sesgo que favorezca a ningún grupo. También se encontró que no hay diferencias significativas en el intercepto de la variable latente, lo que sugiere que la fatiga se experimenta independientemente de la variable sexo.

Limitaciones

Es importante resaltar que en este estudio no se utilizaron pruebas físicas para medir la fatiga y

se consideraron sólo medidas de autoinforme, por lo que sería conveniente hacer estudios que las incluyeran, ya que a pesar de que el tratamiento estadístico riguroso de los datos ayuda a controlar el error en las mediciones, se ha observado que en el caso de autoinformes de salud, las personas pueden exagerar o subestimar los síntomas por tener percepciones alteradas debido a su propio estado (Cohen, 1979).

También es necesario mencionar que la información sobre validez presentada en este estudio es pertinente exclusivamente a la medición de la fatiga en población general de la Ciudad de México. Nuevos estudios de validez serían necesarios para conocer las propiedades psicométricas de la escala obtenida en este estudio en otros contextos, con la finalidad de poder utilizarla en los mismos para fines de investigación o diagnóstico. Aún ante la posibilidad de trabajar con población clínica dentro de la Ciudad de México, serían necesarios nuevos estudios de validez; considerando que la estructura de los reactivos podría ser diferente, y que en la presente versión se han eliminado algunos reactivos.

Se debe destacar que la prueba de invarianza demuestra que la escala es una medida insesgada por sexo, y por consiguiente es adecuado usarla para comparar la fatiga entre hombres y mujeres; sin embargo, en caso de utilizarse para otras variables atributivas, debería establecerse primero que funcione de la misma manera para todos los grupos.

Conclusión

Los resultados encontrados muestran que la FSS tiene propiedades psicométricas que la hacen una medida confiable y válida para el propósito de evaluar fatiga en población general de la ciudad de México y puede ser utilizada en investiga-

ciones o evaluaciones que contribuyan a conocer la prevalencia y las consecuencias de la fatiga en la población general.

También puede contribuir en el diagnóstico, lo que permitiría a los profesionales de la salud atender este problema desde sus etapas tempranas y contribuir al desarrollo de programas de prevención de la fatiga.

Referencias

- Ali, M. R., Khaleque, A., Khanam, M., Al-Shatti, A., & Ahmed, R. U. (1994). Critical flicker frequency of mentally retarded and normal persons. *Perceptual and Motor Skills*, 79(3), 1235-1238. doi: 10.2466/pms.1994.79.3.1235
- Armutlu, K., Keser, İ., Korkmaz, N., Akbıyık, D. İ., Sümbüloğlu, V., Güney, Z., & Karabudak, R. (2007). Psychometric study of Turkish version of Fatigue Impact Scale in multiple sclerosis patients. *Journal of the Neurological Sciences*, 255(1-2), 64-68. doi: 10.1016/j.jns.2007.01.073
- Bernal-Vargas, L., Riveros-Munévar, F., Vinaccia-Alpi, S., & Quiceno-Sierra, J. M. (2017). Estructura factorial y consistencia interna de la Escala de Severidad de Fatiga en población colombiana con enfermedades crónicas. *Enfermería Global*, 16(2), 37. doi: 10.6018/eglobal.16.2.255821
- Beurskens, A. J., Bültmann, U., Kant, I., Vercoulen, J. H., Bleijenberg, G., & Swaen, G. M. (2000). Fatigue among working people: Validity of a questionnaire measure. *Occupational and Environmental Medicine*, 57(5), 353-357. Recuperado de <https://www.ncbi.nlm.nih.gov/pubmed/10769302>
- Chalder, T., Berelowitz, G., Pawlikowska, T., Watts, L., Wessely, S., Wright, D., & Wallace, E. P. (1993). Development of a fatigue scale. *Journal of Psychosomatic Research*, 37(2), 147-53. Recuperado de <https://www.ncbi.nlm.nih.gov/pubmed/8463991>
- Cohen, F. (1979). Personality, stress, and the develop-

- ment of physical illness. En G. C. Stone, F. Cohen, N. E. Adler & Associates. *Health Psychology - A Handbook. Theories, applications, and challenges of a psychological approach to the healthcare system.* (pp. 1-32). San Francisco: JosseyBass.
- De Croon, E. M. (2006). Psychometric properties of the Need for Recovery After Work Scale: Test-retest reliability and sensitivity to detect change. *Occupational and Environmental Medicine*, 63(3), 202-206. doi: 10.1136/oem.2004.018275
- DeShon, R. P. (2004). Measures are not invariant across groups without error variance homogeneity. *Psychology Science*, 46(1), 137-149. Recuperado de http://www.pabst-publishers.de/psychology-science/1-2004/abstract_08.html
- Díaz-Guerrero, R. (2003). *Psicología del Mexicano* (6ª imp.). México: Trillas.
- Duarte-Ayala, R. E. (2017). *Fatiga, estilos de enfrentamiento y calidad de vida en médicos residentes* (Tesis doctoral). Universidad Nacional Autónoma de México, Facultad de Psicología, México. Recuperado de <http://132.248.9.195/ptd2017/junio/513011441/Index.html>
- Duarte-Ayala, R. E., Velasco-Rojano, Á. E., Sánchez-Sosa, J. J., & Reyes-Lagunes, L. I. (2017). Adaptación y validación de la Escala de Impacto de Fatiga. *Acta de Investigación Psicológica*, 7(1), 2585-2592. doi: 10.1016/j.aijpr.2017.02.002
- Duarte, R. E., Velasco, A. E., & Reyes-Lagunes, L. I. (2016). Validación psicométrica de la Escala de Gravedad de Fatiga. En C. Gómez, J. E. Hernández-Sánchez, R. Díaz-Loving, L. I. Reyes-Lagunes & S. Rivera (Eds.), *Aportaciones Actuales de la Psicología Social. Volumen III.* (pp. 103-111) Tabasco, México: Asociación Mexicana de Psicología Social AMEPSO. Universidad Autónoma Juárez de Tabasco.
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, 4(3), 272-299. doi: 10.1037/1082-989x.4.3.272
- Fuentes-Márquez, S., Senín-Calderón, C., Rodríguez-Testal, J. F., & Carrasco, M. A. (2015). Perceived experience of fatigue in clinical and general population: Descriptors and associated reactivities. *Spanish Journal of Psychology*, 18(9), 1-8. doi: 10.1017/sjp.2015.11
- Gök, K., Cengiz, G., Erol, K., & Özgöçmen, S. (2016). The Turkish Version of Multidimensional Assessment of Fatigue and Fatigue Severity Scale is reproducible and correlated with other outcome measures in patients with systemic sclerosis. *Archives of Rheumatology*, 31(4), 329-332. doi: 10.5606/ArchRheumatol.2016.5909
- Gottschalk, M., Kümpfel, T., Flachenecker, P., Uhr, M., Trenkwalder, C., Holsboer, F., & Weber, F. (2005). Fatigue and regulation of the hypothalamo-pituitary-adrenal axis in multiple sclerosis. *Archives of Neurology*, 62(2), 277-280. doi: 10.1001/archneur.62.2.277
- Instituto Nacional de Desórdenes Neurológicos y Accidentes Vasculares [NINDS] (2002). *Esclerosis Múltiple: Esperanza en la Investigación.* Bethesda, Maryland: Departamento de Salud y Servicios Humanos de los Estados Unidos. Recuperado de https://espanol.ninds.nih.gov/trastornos/span_esclerosis.pdf
- Jackson, D. L. (2003). Revisiting sample size and number of parameter estimates: Some support for the N:q hypothesis. *Structural Equation Modeling*, 10(1), 128-141. doi: 10.1207/S15328007SEM1001_6
- Johansson, S., Kottorp, A., Lee, K. A., Gay, C. L., & Lerdal, A. (2014). Can the Fatigue Severity Scale 7-item version be used across different patient populations as a generic fatigue measure –A comparative study using a Rasch model approach. *Health Quality of Life Outcomes*, 12(1), 24. doi: 10.1186/1477-7525-12-24
- Kenny, D. A., & Milan, S. (2012). Identification: A nontechnical discussion of a technical issue. En R. H. Hoyle (Ed.), *Handbook of Structural Equation Modeling* (pp. 145-163). New York: Guilford Press.
- Kos, D., Kerckhofs, E., Carrea, I., Verza, R., Ramos, M., & Jansa, J. (2005). Evaluation of the Modified

- Fatigue Impact Scale in four different European countries. *Multiple Sclerosis*, 11(1), 76-80. doi: 10.1191/1352458505ms1117oa
- Kline, R. (2016). *Principles and practice of structural equation modeling* (4^a ed.). New York, NY: The Guilford Press.
- Krupp, L. B., LaRocca, N. G., Muir-Nash, J., & Steinberg, D. (1989). The Fatigue Severity Scale. Application to patients with multiple sclerosis and systemic lupus erythematosus. *Archives of Neurology*, 46(10), 1121-1123. doi: 10.1001/archneur.1989.00520460115022
- Lerdal, A., & Kottorp, A. (2011). Psychometric properties of the Fatigue Severity Scale—Rasch analyses of individual responses in a Norwegian stroke cohort. *International Journal of Nursing Studies*, 48(10), 1258-1265. doi: 10.1016/j.ijnurstu.2011.02.019
- Lerdal, A., Kottorp, A., Gay, C., Aouizerat, B. E., Portillo, C. J., & Lee, K. A. (2011). A 7-item version of the Fatigue Severity Scale has better psychometric properties among HIV-infected adults: An application of a Rasch model. *Quality of Life Research*, 20(9), 1447-1456. doi: 10.1007/s11136-011-9877-8
- MacCallum, R. C., & Hong, S. (1997). Power analysis in covariance structure modeling using GFI and AGFI. *Multivariate Behavioral Research*, 32(2), 193-210. doi: 10.1207/s15327906mbr3202_5
- MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin*, 111(3), 490-504. Recuperado de <http://www.apa.org/pubs/journals/bul>
- Magnusson, D. (1998). *Teoría de los tests*. México: Trillas.
- Malta-Purim, K. S., Bittencourt-Gumarães, A. T., Kapp-Titski, A. C., & Leite, N. (2016). Sleep deprivation and drowsiness of medical residents and medical students. *A Revista do Colégio de Cirurgiões*, 43(6), 438-444. doi: 10.1590/0100-69912016006005
- Millsap, R. E., & Olivera-Aguilar, M. (2012). Investigating measurement invariance using confirmatory factor analysis. En R. H. Hoyle (Ed.), *Handbook of Structural Equation Modeling* (pp. 380-392). New York, NY: Guilford Press.
- 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.
- Neisa, C., Tangarife, A., & Godoy, L. (2014). Variables moderadoras de la relación fatiga laboral y accidentes de trabajo en personal operativo. En J. Orejuela (Ed.), *Psicología de las Organizaciones y del Trabajo* (pp. 429-446). Cali: Editorial Bonaventuriana.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Teoría psicométrica*. McGraw-Hill: New York.
- Oncu, J., Atamaz, F., Durmaz, B., & On, A. (2013). Psychometric properties of Fatigue Severity and Fatigue Impact Scales in postpolio patients. *International Journal of Rehabilitation Research*, 36(4), 339-345. doi: 10.1097/mrr.0b013e3283646b56
- Orrantia-Daniel, M. C. G., & De la Vega-Bustillos, E. (2003). Trabajo, fatiga, calidad y productividad. *IV Congreso Nacional de Estudios del Trabajo*. Colegio de Sonora. Hermosillo, Sonora. Recuperado de <http://www.izt.uam.mx/sotraem/Documentos/Documentos/Amet2003/templates/res/tema%209/GILBERTO%20ORRANTIA-ENRIQUE%20DE%20LA%20VEGA.pdf>
- Ottoneo, M., Pellicciari, L., Giordano, A., & Foti, C. (2016). Rasch analysis of the Fatigue Severity Scale in Italian subjects with multiple sclerosis. *Journal of Rehabilitation Medicine*, 48(7), 597-603. doi: 10.2340/16501977-2116
- Reyes-Lagunes, L. I. (1993). Las redes semánticas naturales, su conceptualización y su utilización en la construcción de instrumentos. *Revista de Psicología Social y Personalidad*, 9(1), 81-97.
- Reyes-Lagunes, L. I., & García y Barragán, L. F. (2008). Procedimiento de validación psicométrica culturalmente relevante: Un ejemplo. En S. Rivera-Aragón, R. Díaz-Loving, R. Sánchez-Aragón & L. I. Reyes-Lagunes (Eds.), *La Psicología Social en México. Vol. XII* (pp. 625-636). México: Asociación Mexicana de Psicología Social.
- Rosseel, Y. (2012). *Lavaan: An R package for structural*

- equation modeling. *Journal of Statistical Software*, 48(2), 1-36. doi: 10.18637/jss.v048.i02
- Schmitt, T. A. (2011). Current methodological considerations in exploratory and confirmatory factor analysis. *Journal of Psychoeducational Assessment*, 29(4), 304-321. doi: 10.1177/0734282911406653
- Torres-Harding, S. R., Jason, L. A., & Taylor, R. R. (2002). Fatigue severity, attributions, medical utilization, and symptoms in persons with chronic fatigue. *Journal of Behavioral Medicine*, 25(2), 99-113. Recuperado de <https://link.springer.com/article/10.1023/A:1014850819995>
- Torres-Harding, S. R., Mason-Shutter, J., & Jason, L. A. (2008). Fatigue Among Spanish- and English-Speaking Latinos. *Social Work in Public Health*, 23(5), 55-72. doi: 10.1080/19371910802053232
- Useche, S., Cendales, B., & Gómez, V. (2017). Measuring fatigue and its associations with job stress, health and traffic accidents in professional drivers: The case of BRT operators. *EC Neurology*, 4(4), 103-118.
- Van't Leven, M., Zielhuis, G.A., Van der Meer, J. W., Verbeek, A. L., & Bleijenberg, G. (2009). Fatigue and chronic fatigue syndrome-like complaints in the general population. *European Journal of Public Health*, 20(3), 251-257. doi: 10.1093/eurpub/ckp113
- Viner, R., & Christie, D. (2005). Fatigue and somatic symptoms. *British Medical Journal*, 330, 1012-1015. doi: 10.1136/bmj.330.7498.1012
- West, S. G., Taylor, A. B., & Wu, W. (2012). Model fit and model selection in structural equation modeling. En R. H. Hoyle (Ed.), *Handbook of Structural Equation Modeling* (pp. 209-231). New York: Guilford Press.
- Whitehead, L. (2009). The measurement of fatigue in chronic illness: A systematic review of unidimensional and multidimensional fatigue measures. *Journal of Pain and Symptom Management*, 37(1), 107-128. doi: 10.1016/j.jpainsymman.2007.08.019
- Yoshitake, H. (1971). Three characteristic patterns of subjective fatigue symptoms. *Ergonomics*, 21(3), 231-233. doi: 10.1080/00140137108931236

Evidencia de Validez y Confiabilidad de una Medida de Soledad en Adolescentes Peruanos

Evidence of Validity and Reliability of a Measure of Loneliness in Peruvian Adolescents

José Ventura-León *¹, Tomás Caycho-Rodríguez¹, Miguel Barboza-Palomino¹,
Georgina Cáceres-Gonzales²

1 - Universidad Privada del Norte, Perú.

2 - Universidad Inca Garcilaso de la Vega, Perú.

Introducción
Método
Resultados
Discusión
Referencias

Recibido: 14/08/2017 Revisado: 18/09/2017 Aceptado: 27/09/2017

Resumen

En este artículo se analizó la evidencia de validez basada en la estructura interna y la confiabilidad de las puntuaciones de los 11 ítems de la Escala de Soledad De Jong-Gierveld (DJGLS, por sus siglas en inglés) en adolescentes peruanos. Se trabajó con una muestra de 851 adolescentes ($M = 15.07$ años, $DE = 1.5$), estudiantes de colegios de Lima Metropolitana. El análisis factorial mostró una estructura bidimensional oblicua (soledad social y emocional) con buenos índices de bondad de ajuste ($CFI \geq .90$, $RMSEA \leq .08$). La confiabilidad fue estimada mediante el coeficiente omega ($\omega = .827$), y considerada como aceptable ($\geq .70$). Se concluye que la DJGLS muestra propiedades psicométricas adecuadas y puede ser considerada una medida breve para el estudio de la soledad en los adolescentes peruanos. Además, el uso de la DJGLS en base a esta estructura factorial oblicua permitirá ampliar el conocimiento de los factores individuales y sociales que participan en el desarrollo de la soledad.

Palabras Clave: *adolescentes, confiabilidad, soledad, validez*

Abstract

In this article evidence of validity was analyzed, based on the internal structure and reliability of the scores of the 11 items of the De Jong-Gierveld Loneliness Scale (DJGLS) in Peruvian adolescents. The study was carried out with a sample of 851 adolescents ($M = 15.07$ years old, $SD = 1.5$) from schools in Metropolitan Lima. Factor analysis showed a two-dimensional oblique structure (social and emotional loneliness) with good indices of goodness of fit ($IFC \geq .90$, $RMSEA \leq .08$). Reliability was estimated using the omega coefficient ($\omega = .827$), and considered as acceptable ($\geq .70$). It is concluded that the DJGLS shows adequate psychometric properties and can be considered a brief measure for the study of loneliness in Peruvian adolescents. Also, the use of DJGLS with this factorial structure allows a better knowledge of the individual and social factors involved in the development of loneliness.

Keywords: *adolescents, reliability, loneliness, validity*

*Correspondencia a: José Ventura-León. Dirección Postal: Av. Tingo María 1122-Cercado de Lima, Lima, Perú. jose.ventura@upn.pe

Cómo citar este artículo: Ventura-León, J., Caycho-Rodríguez, T., Barboza-Palomino, M., & Cáceres-González, G. (2017). Evidencia de validez y confiabilidad de una medida de soledad en adolescentes peruanos. *Revista Evaluar*, 17(2), 126-142. Recuperado de <https://revistas.unc.edu.ar/index.php/revaluar>

Introducción

La *soledad* es definida como un estado emocional subjetivo, desagradable y estresante; originado por la percepción de que la cantidad y calidad de las relaciones sociales actuales no son las deseadas (Asher & Weeks, 2014; Perlman & Peplau, 1981) y caracterizada por una profunda sensación de aislamiento social, amenaza personal y falta de control (Hawkley & Cacioppo, 2010; VanderWeele, Hawkley, & Cacioppo, 2012). Algunos autores (Rubio & Aleixandre, 2001) señalan una diferencia entre sentirse solo y estar solo, indicando que el primero es un sentimiento subjetivo y doloroso; y el segundo, se entiende como la ausencia de compañía social y familiar. De esta forma, considerando lo anterior, se identifican algunas características importantes del concepto de *soledad*, entre estas: (a) tiene su origen en el establecimiento inadecuado de relaciones significativas; (b) se experimenta de forma subjetiva y es diferente al aislamiento social, que implica el distanciamiento físico; (c) genera una sensación estresante y de desagrado (Perlman & Peplau, 1981).

En los últimos años, la soledad se ha convertido en un problema de salud pública (Holt-Lunstad, Smith, Baker, Harris, & Stephenson, 2015), debido a los efectos que tiene en la salud física y psicológica (Doman & Le Roux, 2010), y que repercuten en la calidad de vida de las personas (Adameczyk & Segrin, 2015). En el plano de la salud física, la soledad es un factor de riesgo de mortalidad tan importante como la obesidad (Holt-Lunstad et al., 2015). Al respecto, un trabajo de meta-análisis (Holt-Lunstad, Smith, & Layton, 2010) concluyó que el riesgo de mortalidad por la soledad puede llegar a ser el doble que el ocasionado por la obesidad. Esta información se encuentra asociada con los problemas que la soledad parece generar en la salud mental, en re-

lación con la depresión (Cacioppo, Hughes, Waite, Hawkley, & Thisted, 2006; Victor & Yang, 2012), los problemas de estrés y el consumo de alcohol (Mushtaq, Shoib, Shah, & Mushtaq, 2014), la angustia (Rokach, 2012), la aparición de fobias y desórdenes obsesivo-compulsivos (Meltzer et al., 2013) y los afectos negativos que funcionan como variable mediadora para la soledad (Baytemir & Yildiz, 2017). Lo anterior evidencia que la soledad puede generar, junto con otros factores, serios problemas en el ámbito físico, cognitivo, emocional, social y comportamental de las personas (Cardona, Villamil, Henao, & Quintero, 2010; Hawkley & Cacioppo, 2010; Holt-Lunstad et al., 2015).

El panorama descrito resulta aún más preocupante si se consideran los datos de prevalencia de la soledad a nivel internacional. Un estudio a gran escala en Estados Unidos señaló que el 79% de personas menores de 18 años informaron sentirse solos (a veces o frecuentemente; Parlee, 1979). A esto se suman los resultados de otro estudio que señaló que los estudiantes de nivel secundario de Estados Unidos reportaron un mayor sentimiento de soledad en comparación con los universitarios (Schultz & Moore, 1988). Esto coincide con el 66% de los escolares estadounidenses que manifestaron haber experimentado soledad en un año (Culp, Clyman, & Culp, 1995), el 50% que manifestó experimentar un sentimiento de soledad recurrente, donde entre el 10% al 20% de éstos describieron la experiencia de soledad como dolorosa (Brennan, 1982). Los resultados expuestos indican que la soledad es un factor de riesgo en la adolescencia (Cava, Musitu, & Murgui, 2007; Hall-Lande, Eisenberg, Christenson, & Neumark-Sztainer, 2007; Heinrich & Gullone, 2006; Woodhouse, Dykas, & Cassidy, 2012) y que está presente en un alto porcentaje de esta población (Qualter et al. 2013).

La aparición de cambios importantes en la adolescencia respecto a los roles, relaciones y ex-

pectativas sociales (Grygiel, Humenny, & Rebisz, 2016; Houghton et al., 2014; Laursen & Hartl, 2013; Qualter et al., 2015), así como la presencia de otras características individuales: rasgos introvertidos, bajos niveles de autoestima (Tiwari, 2013; Vanhalst, Luyckx, & Goossens, 2014), inadecuadas competencias sociales y estrategias de afrontamiento, ausencia de apoyo social (Mahon, Yarcheski, Yarcheski, Cannella, & Hanks, 2006; Tiwari, 2013) entre otros, también constituyen factores antecedentes de la soledad (Madsen, et al., 2016).

Lo expuesto anteriormente caracteriza a la adolescencia como una etapa crítica, donde se hace evidente la posibilidad de que la soledad alcance niveles crónicos e incluso patológicos (Galanaki, Polychronopoulou, & Babalis, 2008; Miller, 2011; Mushtaq et al., 2014), con graves consecuencias para el bienestar de los adolescentes, debido a su asociación con problemas comportamentales, académicos y de salud (Goosby, Bellatorre, Walsemann, & Cheadle, 2013; Van Dulmen & Goossens 2013). En ese sentido, en la población adolescente los sentimientos de soledad aumentan el riesgo de aparición de síntomas depresivos (Ladd & Ettekal, 2013; Mcwhirter, Besett-Alesch, Horibata, & Gat, 2002), trastornos de la alimentación (Herpertz-Dahlmann, 2015; Levine 2012), problemas académicos (Benner & Crosnoe 2011), ideación e intentos de suicidio, (Schinka, Van Dulmen, Bossarte, & Swahn, 2012) y consumo de alcohol (Edgren, Castrén, Jokela, & Salonen, 2016; Krause-Parello, 2008; Page, Dennis, Lindsay, & Merrill, 2011).

Uno de los aspectos más estudiados es la posible existencia de diferencias en los niveles de soledad según el sexo durante la adolescencia. Al respecto, la evidencia no ha sido consistente, ya que algunos estudios señalan que las mujeres presentan mayores niveles de soledad (Vanhalst et al., 2012) y otros indican que los niveles más

altos son experimentados por los hombres (Hoza, Bukowski, & Beery, 2000), mientras que la mayoría reporta que no existen diferencias en la soledad entre ambos (Jobe-Shields, Cohen, & Parra, 2011). De acuerdo a Van Roekel, Scholte, Engels, Goossens y Verhagen (2015), las diferencias en los resultados se explicarían por la ausencia de un análisis de la manera en que ambos sexos perciben su contexto social, lo cual es coherente con lo mencionado por otros autores (De Jong-Gierveld & Van Tilburg, 2006; Madsen et al., 2016). Este es el caso de la escuela, donde los adolescentes pasan gran parte de su tiempo (Weeks & Asher, 2012), y que se convierte en un contexto de influencia para su desarrollo social que puede ser percibido de diferente manera por hombres y mujeres (Weare, 2000).

Las consecuencias que tiene la soledad en la salud física y psicológica de los adolescentes hacen necesario contar con medidas de soledad que brinden interpretaciones válidas y confiables. Sin embargo, los instrumentos más utilizados para medir la soledad en los adolescentes, como el Cuestionario de Soledad e Insatisfacción Social (Loneliness and Social Dissatisfaction Questionnaire; Asher, Hymel, & Renshaw, 1984) y la Escala de Soledad para Niños y Adolescentes (Loneliness and Aloneness Scale for Children and Adolescents; Marcoen, Goossens, & Caes, 1987); presentan la limitación de que sus ítems se centran en examinar la soledad en contextos muy específicos como la escuela y familia, lo que dificulta su uso en estudios referidos a diferentes grupos de edad y otros contextos como el trabajo o el vecindario (Grygiel et al., 2016).

En este sentido, resultarían útiles medidas relativamente breves que permitan una comparación de la soledad entre diversos grupos de edad y que no presenten una variación cultural de los ítems, de modo que fuera posible utilizarlas en diferentes países (Grygiel et al., 2016; Van Tilburg,

Havens, & De Jong-Gierveld, 2004). La Escala de Soledad De Jong-Gierveld (DJGLS, por sus siglas en inglés; De Jong-Gierveld & Kamphuis, 1985) es un instrumento de medida que cumple con estas características y que es muy utilizado en Europa y Asia (Buz & Prieto, 2013; De Jong-Gierveld & Van Tilburg, 2010). Construida inicialmente como una medida unidimensional del nivel de severidad de la soledad en adultos mayores (Tomás, Pinazo-Hernandis, Donio-Bellegarde, & Hontangas, 2017), la DJGLS tiene como fundamento teórico al modelo cognitivo, que entiende la experiencia de soledad como la forma en que las personas perciben, experimentan y evalúan su grado de aislamiento y falta de comunicación (De Jong-Gierveld & Kamphuis, 1985; De Jong-Gierveld, 1987).

La DJGLS ha sido traducida y validada en diferentes idiomas para población adulta mayor (Ayala et al., 2012; Buz & Prieto, 2013; Iecovich, 2013; Leung, De Jong-Gierveld, & Lam, 2008). Diversas investigaciones centradas en el análisis de la estructura interna de la DJGLS ponen en duda su unidimensionalidad (Grygiel, Hummenny, Rebisz, Switaj, & Sikorska, 2012), e indican la existencia de dos dimensiones subyacentes, denominadas soledad social y soledad emocional (Buz & Pérez-Arrechaederra, 2014; Buz, Urchaga, & Polo, 2014; Kunts, Bogaerts, & Winkel, 2010; Shiovitz-Ezra & Ayalon, 2012). Algunos autores (Buz & Prieto, 2013; De Jong-Gierveld & Kamphuis, 1985; De Jong-Gierveld & Van Tilburg, 2010) señalan, con base en estudios en diferentes países, que si bien los sentimientos de soledad son heterogéneos, por lo que generan experiencias sociales y emocionales, la presencia de estas dos dimensiones es atribuible al efecto del método de los ítems inversos, de modo que no existe fundamento teórico para sustentar la presencia de dos dimensiones. Otros estudios (De Jong-Gierveld & Van Tilburg, 1999; Zammuner,

2008) que defienden la unidimensionalidad de la DJGLS, recomiendan asimismo la interpretación bidimensional cuando esta sea adecuada para el tipo de pregunta de investigación.

Si bien la DJGLS ha sido utilizada en diferentes investigaciones para evaluar la soledad en los adolescentes (Błachnio, Przepiorka, Boruch, & Bałakier, 2016; Pontes, Griffiths, & Patrão, 2014; Whitehouse, Durkin, Jaquet, & Ziatas, 2009), son pocos los trabajos que analizan su estructura interna en esta población. En un estudio realizado con adolescentes polacos (Grygiel et al., 2016), se indicó que la soledad evaluada mediante la DJGLS es mejor conceptualizada mediante un modelo bifactor, con un factor general con dos factores ortogonales específicos. A pesar de esto, los autores citados señalan que el 75% de la varianza común explicada por el factor general, los elevados valores de los coeficientes omega y omega H (índice de unidimensionalidad esencial) para el factor general, así como los bajos valores de omega HS (que permite estimar la confiabilidad de las subescalas luego de controlar el factor general) para los factores específicos, indicarían que la DJGLS es esencialmente unidimensional. En Turquía (Çavdar, Bağcı, Çorbacı, Sarıtaş, & Taşdelen-Yayvak, 2015), un estudio llevado a cabo con estudiantes entre los 18 y los 25 años determinó que la DJGLS presenta una estructura bidimensional (soledad emocional y soledad social), con adecuados valores del coeficiente de consistencia interna alfa de Cronbach ($\alpha = .873$).

Considerando las repercusiones de la soledad en las relaciones sociales, salud física y psicológica de los adolescentes, resulta necesario disponer de una escala de medición con adecuadas propiedades psicométricas para la población adolescente peruana. En este sentido, el objetivo del estudio, de tipo instrumental (Ato, López, & Benavente, 2013), es examinar la evidencia de validez basada en la estructura interna y la confia-

bilidad de las puntuaciones de la DJGLS en una población no clínica de adolescentes peruanos.

El estudio se justifica por diversos aspectos: en primer término, la DJGLS ha sido validada en jóvenes y adultos peruanos (Ventura-León & Caycho, 2017); sin embargo, es necesario conocer el comportamiento estadístico de la DJGLS en la población de adolescentes, debido a los cambios cognitivos, físicos, sociales y emocionales que se producen en esta etapa del desarrollo (Laursen & Hartl, 2013), y por la existencia de diferentes características ya expuestas que pueden constituir antecedentes a la presencia de la soledad (Grygiel et al., 2016; Houghton et al., 2014; Laursen & Hartl, 2013; Levitt, Guacci-Franco, & Levitt, 1993; Madsen, et al., 2016; Mahon et al., 2006; Qualter et al., 2015; Tiwari, 2013; Vanhalst et al., 2014; Woodhouse et al., 2012). En segundo lugar, la evidencia señalada anteriormente analizada revela que un porcentaje alto de adolescentes experimentan el sentimiento de soledad (Brennan, 1982; Culp et al., 1995; Parlee, 1979), no obstante en Latinoamérica y específicamente en el Perú, no se cuenta con cifras provenientes de instrumentos validados de autoinforme. En tercer lugar, la DJGLS recientemente ha sido validada en la población de adolescentes en Polonia (Grygiel et al., 2016); empero, no existe evidencia de validez en países latinoamericanos, por lo que los resultados que se informan ayudarán a una mejor comprensión de la soledad en el contexto peruano y latinoamericano. Cuarto, por ser la DJGLS una medida breve de autoinforme, sería útil en investigaciones epidemiológicas y en la evaluación del impacto de intervenciones.

Método

Participantes

La muestra estuvo conformada por 851 adolescentes, 417 mujeres (49%) y 434 varones (51%), con una edad promedio de 15.07 años ($DE = 1.5$), quienes cursaban el primer (9%), segundo (8.2%), tercer (7.6%), cuarto (37.7%) y quinto (37.4%) grado de educación secundaria en colegios de Lima Metropolitana. Respecto al lugar de nacimiento, 698 (82%) habían nacido en Lima, mientras que los 153 restantes (18%) en provincia. El análisis de comparación de medias no evidenció diferencias significativas en el promedio de edad en función al sexo ($t_{(497)} = .884, p = .337$).

Instrumentos

Se creó un cuestionario sociodemográfico ad hoc, destinado a recabar información acerca de la edad, sexo, grado o sección educativa de nivel secundario y lugar de nacimiento.

Asimismo, se utilizó la versión de 11 ítems de la Escala de Soledad de De Jong-Gierveld (DJGLS; De Jong-Gierveld & Kamphuis, 1985), validada para jóvenes y adultos peruanos por Ventura-León y Caycho (2017). La misma cuenta con una alta confiabilidad ($\omega = .83$) y brinda además, evidencia de una estructura de dos factores relacionados con adecuadas bondades de ajuste. Esta versión fue adaptada lingüísticamente a la población adolescente mediante una muestra piloto que buscaba garantizar la claridad de los reactivos para dicha población. Así, se solicitó a 10 estudiantes de los grados a evaluarse que leyeran los ítems y respondan a las siguientes preguntas: *¿Se entiende?*, *¿Es claro?*, *¿Puedes decir lo mismo de una manera más fácil?* La Tabla 1 presenta los ítems originales y la versión final luego de los cambios sugeridos.

Tabla 1

Diferencias entre algunos ítems de la DJGLS de Ventura-León y Caycho (2017) y los utilizados en este estudio.

Nº ítem	Versión de Ventura-León y Caycho (2017) para jóvenes y adultos	Ítems modificados para el presente estudio en adolescentes
1	Siempre hay alguien con quien puedo hablar de mis problemas diarios.	Siempre hay alguna persona con quien puedo hablar de mis problemas.
2	Echo de menos tener un buen amigo de verdad.	Extraño tener un(a) buen(a) amigo(a) verdadero(a).
4	Hay suficientes personas a las que puedo recurrir en caso de dificultades.	Hay suficientes personas a las que puedo buscar en caso de problemas.
5	Echo de menos la compañía de otras personas.	Extraño la compañía de otras personas.
6	Pienso que mi círculo de amistades es demasiado limitado.	Pienso que la cantidad de amigos(as) que tengo son pocos.
7	Tengo mucha gente en la que puedo confiar completamente.	Tengo muchas personas en las que puedo confiar mucho.
9	Echo de menos tener gente a mi alrededor.	Extraño tener personas a mi alrededor.
10	Me siento abandonado(a) frecuentemente.	Me siento solo(a) casi siempre.
11	Puedo contar con mis amigos siempre que lo necesito.	Puedo contar con mis amigos(as) siempre que lo necesito.

La DJGLS tiene tres categorías de respuesta (1 = *no*, 2 = *más o menos*, 3 = *sí*). Siguiendo las recomendaciones para su calificación (De Jong-Gierveld & Kamphuis, 1985), se dicotomizaron las respuestas otorgando un punto si la respuesta era *más o menos* o *no* en los ítems: 1, 4, 7, 8 y 11 (ítems inversos) y un punto a las respuestas *más o menos* o *si* en los ítems: 2, 3, 5, 6, 9 y 10 (ítems directos). Una vez dicotomizada la escala, los puntajes pueden variar desde un valor mínimo de cero (nula soledad) a un máximo de 11 (soledad extrema), denotando el continuo de soledad.

Procedimiento

Inicialmente, se realizó una entrevista semiestructurada con diez adolescentes para comprobar la claridad de los ítems; así, se modificaron levemente algunos términos con la finalidad de aumentar la comprensión de los reactivos. En segundo lugar, tras la revisión se obtuvo el aval,

con respecto a los aspectos éticos y metodológicos, de la universidad de origen de los autores del presente estudio. En tercer lugar, se solicitaron los permisos correspondientes a las autoridades de las instituciones educativas para la aplicación de la DJGLS, informándoles acerca del objetivo y el alcance de la investigación. Conseguida la autorización, y antes de la aplicación, se informó el objetivo de la investigación a los estudiantes y a sus padres, y se obtuvo la participación voluntaria mediante la firma del consentimiento informado de estos últimos. La DJGLS fue administrada de manera colectiva en un tiempo aproximado de 15 minutos.

Análisis de datos

El análisis estadístico se efectuó con el apoyo del programa computacional de acceso libre “R” versión 3.1.2 (R Development Core Team, 2007). En ese sentido, se dividió el proceso en

dos etapas: Primero, se examinaron los estadísticos descriptivos de los ítems: media, desviación estándar, curtosis y asimetría (Hair, Anderson, Tatham, & Black, 2005). Segundo, se corroboró la estructura factorial mediante un análisis factorial confirmatorio (AFC) utilizándose los índices de ajuste recomendados (Hooper, Coughlan, & Mullen, 2008): chi-cuadrada con corrección Satorra-Bentler (S-B; Satorra & Bentler, 2001), la razón entre chi-cuadrada y los grados de libertad [$S-B\chi^2/df$], según la cual valores menores o iguales a 2 indican un buen ajuste (Tabachnick & Fidell, 2007); los índices de ajuste absoluto (Hu & Bentler, 1999): error de aproximación de la media cuadrática [RMSEA \leq .06], raíz cuadrada de la media cuadrática del residual estandarizado [SRMR \leq .08], y el índice comparativo (Hu & Bentler, 1999): índice de ajuste comparativo [CFI \geq .90]. Adicionalmente, se estimó el criterio de información de Akaike (CIA; Caballero, 2011). Cabe señalar que todos estos índices fueron calculados mediante métodos robustos.

En vista de que se consideró un modelo bifactor y con efecto del método, se tomaron en cuenta las medidas auxiliares que evalúan la dimensionalidad de modelos bifactor (Reise, Moore, & Haviland, 2010; Reise, Scheines, Widaman, & Haviland, 2013; Rios & Wells, 2014; Rodriguez, Reise, & Haviland, 2016): *Explained Common Variance General* (ECV_{Gen}); *Explained Common Variance Specific* ($ECV_{Specific}$); *Individual Explained Common Variance* ($IECV_{Gen}$); *Percentage of Uncontaminated Correlations* (PUC); y la revisión de las cargas de factor general y específico: *Average Relative Parameter Bias* (ARPB). Asimismo, se probó la confiabilidad del modelo bifactor mediante: coeficiente omega (ω ; McDonald, 1999), omega de subescala (ω_s), omega jerárquico (ω_H ; McDonald, 1999; Zinbarg, Barlow, & Brown, 1997), y omega jerárquico de subescala (ω_{HS} ; Reise et al., 2013).

Resultados

Análisis preliminar de los ítems

En la Tabla 2 se presenta el análisis preliminar de los ítems. En relación con la media aritmética, se observó que el ítem 2 (*Extraño tener un/a buen/a amigo/a verdadero/a*), evidenciaba el mayor valor, mientras que el ítem 1 (*Siempre hay alguna persona con quien puedo hablar de mis problemas*), mostraba el menor valor. Las desviaciones estándar de los ítems 7, 3, 6 y 9 fueron las más altas, lo que indicaba mayor heterogeneidad en las respuestas de los participantes. Por su parte, los ítems 2, 5, 7 y 9 presentaron asimetría negativa evidenciando así una tendencia de los participantes a responder con valores bajos en la escala, mientras que con el resto de los ítems sucedió lo contrario. Todos los valores de curtosis observados fueron inferiores a cero, lo que indica una baja concentración de valores hacia el centro de la distribución.

Análisis factorial confirmatorio

Se realizó un AFC con el fin de someter los hallazgos previos a un método más robusto. Con este propósito se modelaron seis estructuras factoriales: (a) Modelo 1, un solo factor con 11 ítems; (b) Modelo 2, dos factores ortogonales (soledad emocional y soledad social); (c) Modelo 3, dos factores oblicuos (soledad emocional y soledad social); (d) Modelo 4, un factor general con dos factores específicos (soledad emocional y soledad social), llamado también modelo bifactor (Holzinger & Swineford, 1937); (f) Modelo 5, un factor general con efecto del método del Factor 1 (ítems directos); y (g) Modelo 6, un factor general con efecto del método del Factor 2 (ítems inversos). Se calculó el coeficiente de Mardia (1970) que dio como resultado un valor de 135.469,

Tabla 2

Análisis preliminar de los ítems (N = 851).

Ítem	M	σ	g1	g2
1. Siempre hay alguna persona con quien puedo hablar de mis problemas.	.25	.44	1.14	-.70
2. Extraño tener un(a) buen(a) amigo(a) verdadero(a).	.73	.44	-1.06	-.87
3. Siento una sensación de vacío a mi alrededor.	.48	.50	.08	-1.99
4. Hay suficientes personas a las que puedo buscar en caso de problemas.	.39	.49	.44	-1.80
5. Extraño la compañía de otras personas.	.72	.45	-.99	-1.03
6. Pienso que la cantidad de amigos(as) que tengo son pocos.	.47	.50	.14	-1.98
7. Tengo varias personas en las que puedo confiar mucho.	.51	.50	-.03	-1.99
8. Hay suficientes personas con las que tengo una amistad muy cercana.	.29	.46	.91	-1.16
9. Extraño tener personas a mi alrededor.	.54	.50	-.16	-1.97
10. Me siento solo(a) casi siempre.	.39	.49	.44	-1.80
11. Puedo contar con mis amigos(as) siempre que lo necesito.	.33	.47	.74	-1.44

Nota. M = Media; σ = Desviación estándar; g1 = Asimetría; g2 = Curtosis.

incumpliendo con el supuesto de normalidad multivariada ($> .70$; [Rodríguez & Ruiz, 2008](#)). Por ende, se decidió atenuar los datos mediante la chi-cuadrada de Satorra-Bentler (S-B; [Satorra & Bentler, 2001](#)).

En la Tabla 3 se exponen los estadísticos de bondad de ajuste para cada uno de los modelos señalados. Se observa que el Modelo 1 presentó los peores ajustes. Los modelos 1, 2, 3, 4 y 6 exhibieron ratio S-B $\chi^2/g1 > 2$; de ellos, el Modelo 2 presenta un valor $\geq .08$ en su SRMR (SRMR = .125); y los Modelos 1 (RMSEA = .099) y 2 (RMSEA = .085) obtuvieron valores RMSEA $\geq .06$. Por otro lado, los Modelos 4 y 5 ostentaron valores de ajuste excepcionales. Principalmente, el Modelo 4 presentó valores de bondad de ajuste óptimos (S-B χ^2 (31) = 58.225; $\chi^2/g1$ = 1.878; CFI = .985; RMSEA = .032 [.019, .045]; SRMR = .029; AIC = -3.775). Ante estos resultados, se procedió a estimar la dimensionalidad para modelo bifactor. A continuación, se presentan los índices auxiliares del modelo bifactor para determinar la dimensionalidad de la escala y confiabilidad de las puntuaciones.

Dimensionalidad de la DJGLS

En la Tabla 4 se presenta el índice ECV_{Gen} (.484), que sugirió que la varianza de los ítems explicaba al factor general en 48.4%; lo que permitiría descartar el supuesto de unidimensionalidad ($\leq .85$; [Stucky, Edelen, Vaughan, Tucker, & Butler, 2014](#)). Por otro lado, los valores de $ECV_{Specific1}$ (.230), $ECV_{Specific2}$ (.286) indicaron que los ítems explicaban el 23% y 28.6% de la varianza en los factores específicos 1 y 2 respectivamente. De acuerdo con el análisis del I-ECV sólo los ítems 3, 6 y 10 eran esencialmente unidimensionales, el resto se caracterizaba por medir un factor específico más un factor general. Asimismo, se obtuvo un valor de PUC de .545 que indicaba un mayor sesgo en los coeficientes de estructura y por tanto una tendencia hacia la multidimensionalidad ($< .80$, [Reise et al., 2013](#)). La revisión de cargas factoriales reveló que ocho ítems presentaban un mayor peso en los factores específicos y sólo tres presentaban un peso mayor en el factor general. El ARPB obtenido fue de 35.7%, valor por encima del criterio recomendado (10-15%;

Tabla 3

Índices de ajuste estadístico de 6 modelos de la Escala de Soledad (N = 851).

	S-B χ^2 (gl)*	S-B χ^2/df	SRMR	CFI	AIC	RMSEA [IC 90%]
Modelo 1	408.969 (44)	9.295	.079	.794	320.969	.099 [.090, .108]
Modelo 2	314.665 (44)	7.151	.125	.848	226.665	.085 [.076, .094]
Modelo 3	166.766 (43)	3.878	.052	.930	80.766	.058 [.049, .068]
Modelo 4	58.225 (31)	1.878	.029	.985	-3.775	.032 [.019, .045]
Modelo 5	106.964 (37)	2.890	.034	.961	32.964	.047 [.037, .058]
Modelo 6	152.775 (38)	4.020	.050	.935	76.775	.060 [.050, .069]

Nota. S-B χ^2 = Chi-cuadrada con atenuación de Satorra-Bentler; *gl* = grados de libertad, SRMR = Raíz residual estandarizada cuadrática media, RMSEA = Error de aproximación de la media cuadrática, CFI = Índice de ajuste comparativo, AIC = Criterio de Akaike, IC 90% = Intervalos de confianza al 90%. En *kursivas* = modelos entre los cuales se tomó la decisión final.

Muthén, Kaplan, & Hollis, 1987), lo que indicó que las cargas factoriales de los modelo bifactor y unidimensional tendían a ser diferentes. Estos resultados sugieren que es mejor conceptualizar la DJGLS como un instrumento bidimensional (Modelo 3), cuyos valores de bondad de ajuste son aceptables.

Confiabilidad

En la Tabla 3 se presentan los valores de confiabilidad para todos los modelos mediante el coeficiente omega (ω ; McDonald, 1999) que osciló entre .759 y .827. El valor para omega de las subescalas (ω_s) osciló entre .595 y .672. El valor de omega jerárquico (ω_H ; McDonald, 1999; Zinbarg et al., 1997) fue de .519, lo que revela el bajo grado en que la puntuación total refleja el factor general ($> .75$, Reise et al., 2013). Respecto a los coeficientes omega jerárquico del factor específico, se obtuvieron valores de .161 y .245 para los factores 1 (ω_{h1}) y 2 (ω_{h2}) respectivamente,

que pueden ser considerados bajos (Reise et al., 2013). Esto permite concluir que gran parte de la varianza de la subescala es debida al factor general y no es exclusiva de los factores específicos (Grygiel et al., 2016).

Discusión

La adolescencia es una etapa en la que se experimentan cambios físicos, sociales, cognitivos y emocionales (Laurson & Hartl, 2013) caracterizados por una fuerte tendencia al rechazo o la búsqueda de aceptación (Woodhouse et al., 2012). Asimismo, existe evidencia científica que señala las consecuencias en la salud física y psicológica de la soledad durante la adolescencia (Doman & Le Roux, 2010; Lasgaard, Goossens, & Elklit, 2011) y cifras estadísticas que revelan el alto porcentaje de adolescentes que experimentan la soledad como algo doloroso (Brennan, 1982; Culp et al., 1995; Heinrich & Gullone, 2006; Schultz & Moore, 1988).

Tabla 4
Saturaciones de los ítems en los diversos modelos de la ESDJG (N = 851).

Ítems	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6	I-ECV
1	.493	.555	.544	.320/.451	.550	.253/.499	.335
2	.331	.363	.366	.278/.284	.243/.311	.368	.489
3	.523	.649	.667	.645.166	.409/.459	.663	.938*
4	.538	.583	.580	.348/.470	.575	.312/.496	.354
5	.242	.438	.391	.165/.806	.085/.532	.394	.040
6	.413	.436	.458	.467/.069	.336/.285	.459	.979*
7	.576	.640	.630	.338/.548	.620	.321/.552	.276
8	.598	.610	.621	.425/.446	.624	.385/.487	.476
9	.277	.483	.436	.306/.403	.117/.560	.435	.366
10	.525	.639	.660	.666/.130	.413/.449	.660	.963*
11	.636	.660	.671	.436/.497	.665	.408/.523	.435
ω	.759	.829	.827	.767	.792	.789	-
ω_{F1}	-	-	-	.595	.645	-	-
ω_{F2}	-	-	-	.635	-	.634	-
ω_h	-	-	-	.519	.603	.606	-
ω_{h1}	-	-	-	.161	.189	-	-
ω_{h2}	-	-	-	.245	-	.183	-
ECV _{VGen}	-	-	-	.484	.667	.621	-
ECV _{Specific1}	-	-	-	.230	.333	-	-
ECV _{Specific2}	-	-	-	.286	-	.379	-
PUC	-	-	-	.545	-	-	-
ARPB	-	-	-	.357	-	-	-

Nota. / = simboliza las cargas factoriales del efecto del método; ECV_{Gen} = Explained common variance general; ECV_{Specific} = Explained common variance specific; I-ECV = Individual explained common variance; PUC = Percentage of uncontaminated correlations; ARPB = Average relative parameter bias; ω = Coeficiente omega; ω_s = Omega de subescala; ω_H = Omega jerárquico; ω_{HS} = Omega jerárquico de subescala. En cursivas = Cargas factoriales < .30

En ese contexto, resulta necesario adaptar y validar un instrumento de soledad en adolescentes; entendiendo que un constructo no necesariamente se expresa de igual forma en un mismo contexto y/o población (Fernández, Pérez, Alderete, Richaud, & Liporace, 2011). En este sentido, el objetivo del presente estudio fue proporcionar evidencia acerca de la estructura interna y confia-

bilidad de las puntuaciones de la DJGLS en una muestra de adolescentes de Lima (Perú) mediante AFC y métodos de consistencia interna.

Los principales hallazgos radican en que la DJGLS cuenta con una estructura compleja, al tiempo que el Modelo 4 (bifactor) obtuvo los mejores valores de bondad de ajuste, situación que es similar a lo encontrado por Grygiel et al.,

(2016). En consecuencia, fue necesario examinar la dimensionalidad y confiabilidad de las puntuaciones a través de los índices auxiliares planteados para este tipo de estructura factorial (Reise et al., 2010; Reise et al., 2013; Rios & Wells, 2014; Rodríguez et al., 2016). De ese modo, se concluye que la DJGLS no es esencialmente unidimensional ya que presenta valores $PUC < .80$ ($PUC = .545$), $ECV_{Gen} < .60$ ($ECV_{Gen} = .484$) y $\omega_H < .70$ ($\omega_H = .519$; Reise et al., 2010; Rodríguez et al., 2016). Además, en la revisión de las cargas factoriales, los factores específicos presentaron un mayor peso en comparación con el factor general, lo que lleva a interpretar que los factores específicos en el modelo ejercen una mayor influencia que el factor general. Estos resultados sugieren que la DJGLS debe ser interpretada como una escala bidimensional oblicua (Modelo 3). A pesar de que los valores de bondad de ajuste no son satisfactorios para este modelo, pueden considerarse aceptables ($CFI \geq .90$, $RMSEA \leq .08$; Hu & Bentler, 1999). De esta forma, los investigadores y diferentes profesionales de la salud, que utilicen la DJGLS para medir la soledad, deben considerar que la escala mide dos dimensiones (social y emocional) relacionadas y no una sola. Lo anterior concuerda con estudios en población adolescente (Çavdar et al., 2015) y en otros grupos etarios (Buz et al., 2014; Buz & Pérez-Arrechaederra, 2014; De Jong-Gierveld & Van Tilburg, 1992; Kunts et al., 2010; Shiovitz-Ezra & Ayalon, 2012; Ventura-León & Caycho, 2017; Zammuner, 2008). En ese sentido, distinguir entre la soledad social y la emocional a nivel conceptual permitirá una comprensión de los factores personales propios de la adolescencia de manera independiente (Goosby et al., 2013; Tiwari, 2013; Van Dulmen & Goossens 2013; Vanhalst et al., 2014), tanto a nivel intra como interpersonal.

Por otro lado, la confiabilidad fue estimada mediante el coeficiente ω (McDonald, 1999)

y no con el coeficiente alfa de Cronbach. En este caso el Modelo 3 presentó un valor de .827 que puede ser considerado aceptable ($\omega \geq .70$). Se prefirió este coeficiente por ser el más indicado cuando se utilizan modelos factoriales (Ventura-León & Caycho-Rodríguez, 2017) y se utilizó el coeficiente omega jerárquico (McDonald, 1999) que es ideal para modelos como el bifactor (Reise et al., 2013), a diferencia del coeficiente alfa de Cronbach que podría verse afectado por la varianza de los factores específicos al no medir directamente un único factor (Huysamen, 2006). Los resultados indicaron que la versión peruana de la DJGLS para adolescentes presentaba un nivel aceptable de confiabilidad bajo el método de consistencia interna, interpretación que se relaciona con lo informado en países como: China (Leung et al., 2008), Israel (Iecovich, 2013), Turquía (Çavdar et al., 2015), España (Ayala et al., 2012; Buz & Prieto, 2013) y otros países europeos (De Jong-Gierveld & Van Tilburg, 2010), aunque estos estudios se realizaron en poblaciones diferentes a la adolescente.

Existen ciertas limitaciones en el estudio. En primer lugar, la técnica de muestreo fue no probabilística, lo que podría generar sesgos de muestreo (Lorh, 2000) y afectar la generalización de los hallazgos. Ante esto, próximas investigaciones deberían considerar muestras más representativas de la población de estudio. En segundo lugar, sólo se ha analizado la evidencia de validez basada en el constructo, por lo que se recomienda considerar otras fuentes de validez para futuros estudios, como la relación con otras variables y la validez de contenido (Ventura-León, 2017). Así también, se sugiere revisar el funcionamiento de la escala desde el modelo de Rasch, a partir del cual fue originalmente diseñada. Por último, no se estimó la confiabilidad test-retest ya que el estudio fue de naturaleza transversal, empero cabe señalar que esta estimación es necesaria para ve-

rificar la estabilidad de las puntuaciones (Muñiz, 1994). Por todo lo expuesto anteriormente, sería recomendable continuar con el estudio de las propiedades psicométricas de la DJGLS, especialmente en contextos culturales distintos del contexto de origen de dicho instrumento (Buz et al., 2014).

A pesar de las limitaciones, los resultados sugieren que la DJGLS es una medida breve que ofrece interpretaciones válidas y puntuaciones confiables acerca de la soledad en adolescentes peruanos, lo que resulta de interés para futuros estudios centrados en las variaciones y los factores determinantes de la soledad en esta población.

Referencias

- Adamczyk, K., & Segrin, C. (2015). Direct and indirect effects of young adults' relationship status on life satisfaction through loneliness and perceived social support. *Psychologica Belgica*, 55(4), 196-211. doi: 10.5334/pb.bn
- Asher, S. R., & Weeks, M. S. (2014). Loneliness and belongingness in the college years. En R. J. Coplan & J. C. Bowker (Eds.), *The Handbook of Solitude: Psychological perspectives on social isolation, social withdrawal, and being alone* (pp. 283-301). Hoboken, NJ: John Wiley.
- Asher, S. R., Hymel, S., & Renshaw, P. D. (1984). Loneliness in children. *Child Development*, 55(4), 1456-1464.
- Ato, M., López, J. J., & Benavente, A. (2013). Un sistema de clasificación de los diseños de investigación en psicología. *Anales de Psicología*, 29(3), 1038-1059.
- Ayala, A., Rodríguez-Blázquez, C., Frades-Payo, B., João-Forjaz, M., Martínez-Martín, P., Fernández-Mayoralas, G., & Rojo-Pérez, F. (2012). Psychometric properties of the Functional Social Support Questionnaire and the Loneliness Scale in non-institutionalized older adults in Spain. *Gaceta Sanitaria*, 26(4), 317-324. doi: 10.1016/j.gaceta.2011.08.009
- Baytemir, K., & Yildiz, M. A. (2017). Multiple mediation of loneliness and negative emotions in the relationship between adolescents' social anxiety and depressive symptoms. *Anales de Psicología/Annals of Psychology*, 33(3), 612-620.
- Benner, A. D., & Crosnoe, R. (2011). The racial/ethnic composition of elementary schools and young children's academic and socioemotional functioning. *American Educational Research Journal*, 48(3), 621-646. doi: 10.3102/0002831210384838
- Błachnio, A., Przepiorka, A., Boruch, W., & Bałakier, E. (2016). Self-presentation styles, privacy, and loneliness as predictors of Facebook use in young people. *Personality and Individual Differences*, 94, 26-31. doi: 10.1016/j.paid.2015.12.051
- Brennan, T. (1982). Loneliness at adolescence. En L. A. Peplau & D. Perlman (Eds.), *Loneliness: A sourcebook of current theory, research and therapy* (pp. 269-290). New York: Wiley.
- Buz, J., Pérez-Arrechaederra, D. (2014). Psychometric properties and measurement invariance of the Spanish version of the 11-item De Jong-Gierveld Loneliness Scale. *International Psychogeriatrics* 26(9), 1553-1564. doi: 10.1017/S1041610214000507
- Buz, J., & Prieto, G. (2013). Analysis of the De Jong-Gierveld Loneliness Scale using the Rasch model. *Universitas Psychologica*, 12(3), 971-981. doi: 10.11144/Javeriana.UPSY12-3.aesd
- Buz, J., Urchaga, D., & Polo, M. E. (2014). Factor structure of the De Jong-Gierveld Loneliness Scale in Spanish elderly adults. *Anales de Psicología*, 30(2), 588-596. doi: 10.6018/analesps.30.2.148371
- Caballero, F. F. (2011). *Selección de modelos mediante criterios de información en análisis factorial. Aspectos teóricos y computacionales* (Tesis doctoral). Universidad de Granada, Granada, España.
- Cacioppo, J. T., Hughes, M. E., Waite, L. J., Hawkley, L. C., & Thisted, R. A. (2006). Loneliness as a specific risk factor for depressive symptoms: Cross-sectional and longitudinal analyses. *Psychology and Aging*, 21(1),

- 140-151. doi: 10.1037/0882-7974.21.1.140
- Cardona, J. L., Villamil, M. M., Henao, E., & Quintero, A. (2010). Validación de la escala ESTE para medir la soledad de la población adulta. *Investigación y Educación en Enfermería*, 28(3), 416-427.
- Cava, M. J., Musitu, G., & Murgui, S. (2007). Individual and social risk factors related to overt victimization in a sample of Spanish adolescents. *Psychological Reports*, 101(1), 275-290.
- Çavdar, D., Bağcı, V., Çorbacı, E. C., Sarıtaş, S., & Taşdelen-Yayvak, M. K. (2015). Adaptation of De Jong-Gierveld Loneliness Scale into Turkish. *International Journal of Humanities and Social Science*, 5(9), 84-92.
- Culp, A. M., Clyman, M. M., & Culp, R. E. (1995). Adolescent depressed mood, reports of suicide attempts, and asking for help. *Adolescence*, 30, 827-837.
- De Jong-Gierveld, J. (1987). Developing and testing a model of loneliness. *Journal of Personality and Social Psychology*, 53(1), 119-128. doi: 10.1037/0022-3514.53.1.119
- De Jong-Gierveld, J., & Kamphuis, F. H. (1985). The development of a Rasch-type loneliness scale. *Applied Psychological Measurement*, 9(3), 289-299. doi: 10.1177/014662168500900307
- De Jong-Gierveld, J., & Van Tilburg, T. (1999). Living arrangements of older adults in the Netherlands and Italy: Coresidence values and behaviour and their consequences for loneliness. *Journal of Cross-cultural Gerontology*, 14(1), 1-24.
- De Jong-Gierveld, J., & Van Tilburg, T. (2006). A 6-item scale for overall, emotional, and social loneliness: Confirmatory tests on survey data. *Research on Aging*, 28(5), 582-598. doi: 10.1177/0164027506289723
- De Jong-Gierveld, J., & Van Tilburg, T. (2010). The De Jong-Gierveld short scales for emotional and social loneliness: Tested on data from 7 countries in the UN generations and gender surveys. *European Journal of Ageing*, 7(2), 121-130. doi: 10.1007/s10433-010-0144-6
- De Jong-Gierveld, J., & Van Tilburg, T. G. (1992). Triangulatie in operationaliseringsmethoden. En G. J. N. Bruinsma & M. A. Zwanenburg (Eds.), *Methodologie voor Bestuurskundigen: Stromingen en Methoden* (pp. 273-298). Muiderberg: Coutinho.
- Doman, L. C. H., & Le Roux, A. (2010). The causes of loneliness and the factors that contribute towards it: A literature review. *Tydskrif Vir Geesteswetenskappe*, 50(2), 216-228.
- Edgren, R., Castrén, S., Jokela, M., & Salonen, A. H. (2016). At-risk and problem gambling among Finnish youth: The examination of risky alcohol consumption, tobacco smoking, mental health and loneliness as gender-specific correlates. *Nordic Studies on Alcohol and Drugs*, 33(1), 61-80. doi: 10.1515/nsad-2016-0005
- Fernández, A., Pérez, E., Alderete, A. M., Richaud, M. C., & Liporace, M. F. (2011). ¿Construir o adaptar tests psicológicos? Diferentes respuestas a una cuestión controvertida. *Evaluar*, 10, 60-74. Recuperado de <https://revistas.unc.edu.ar/index.php/revaluar>
- Galanaki, E. P., Polychronopoulou, S. A., & Babalis, T. K. (2008). Loneliness and social dissatisfaction among behaviourally at-risk children. *School Psychology International*, 29(2), 214-229. doi: 10.1177/0143034308090061
- Goosby, B. J., Bellatorre, A., Walsemann, K. M., & Cheadle, J. E. (2013). Adolescent loneliness and health in early adulthood. *Sociological Inquiry*, 83(4), 505-536. doi: 10.1111/soin.12018
- Grygiel, P., Humenny, G., & Rebisz, S. (2016). Using the De Jong-Gierveld Loneliness Scale with early adolescents: Factor structure, reliability, stability, and external validity. *Assessment*. doi: 10.1177/1073191116682298
- Grygiel, P., Humenny, G., Rebisz, S., Switaj, P., & Sikorska, J. (2012). Validating the Polish adaptation of the 11-item De Jong-Gierveld Loneliness Scale. *European Journal of Psychological Assessment*, 29, 129-139. doi: 10.1027/1015-5759/a000130
- Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (2005). *Análisis multivariante*. Madrid: Pearson.

- Hall-Lande, J. A., Eisenberg, M. E., Christenson, S. L., & Neumark-Sztainer, D. (2007). Social isolation, psychological health, and protective factors in adolescence. *Adolescence*, 42(166), 265-286.
- Hawkley, L. C., & Cacioppo, J. T. (2010). Loneliness matters: A theoretical and empirical review of consequences and mechanisms. *Annals of Behavioral Medicine*, 40(2), 218-227. doi: 10.1007/s12160-010-9210-8
- Heinrich, L. M., & Gullone, E. (2006). The clinical significance of loneliness: A literature review. *Clinical Psychology Review*, 26(6), 695-718. doi: 10.1016/j.cpr.2006.04.002
- Herpertz-Dahlmann, B. (2015). Adolescent eating disorders: Update on definitions, symptomatology, epidemiology, and comorbidity. *Child and Adolescent Psychiatric Clinics of North America*, 24(1), 177-196. doi: 10.1016/j.chc.2014.08.003
- Holt-Lunstad, J., Smith, T. B., Baker, M., Harris, T., & Stephenson, D. (2015). Loneliness and social isolation as risk factors for mortality: A meta-analytic review. *Perspectives on Psychological Science*, 10(2), 227-237. doi: 10.1177/1745691614568352
- Holt-Lunstad, J., Smith, T. B., & Layton, J. B. (2010). Social relationships and mortality risk: A meta-analytic review. *PLoS Medicine*, 7(7), 1-20. doi: 10.1371/journal.pmed.1000316
- Holzinger, K. J., & Swineford, F. (1937). The bi-factor method. *Psychometrika*, 2, 41-54.
- Hooper, D., Coughlan, J., & Mullen, M. R. (2008). Structural equation modelling: Guidelines for determining model fit. *The Electronic Journal of Business Research Methods*, 6(1), 53-60.
- Houghton, S., Hattie, J., Wood, L., Carroll, A., Martin, K., & Tan, C. (2014). Conceptualising loneliness in adolescents: Development and validation of a self-report instrument. *Child Psychiatry & Human Development*, 45(5), 604-616. doi: 10.1007/s10578-013-0429-z
- Hoza, B., Bukowski, W. M., & Beery, S. (2000). Assessing peer network and dyadic loneliness. *Journal of Clinical Child Psychology*, 29(1), 119-128. doi: 10.1207/S15374424jccp2901_12
- Hu, L.T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1-55. doi: 10.1080/10705519909540118
- Huysamen, G. K. (2006). Coefficient alpha: Unnecessarily ambiguous; unduly ubiquitous. *SA Journal of Industrial Psychology*, 32(4), 34-40.
- Iecovich, E. (2013). Psychometric properties of the Hebrew version of the De Jong-Gierveld Loneliness Scale. *Educational Gerontology*, 39(1), 12-27. doi: 10.1080/03601277.2012.660860
- Jobe-Shields, L., Cohen, R., & Parra, G. R. (2011). Patterns of change in children's loneliness: Trajectories from third through fifth grades. *Merrill-Palmer Quarterly*, 57(1), 25-47.
- Krause-Parello, C. A. (2008). Loneliness in the school setting. *The Journal of School Nursing*, 24(2), 66-70. doi: 10.1177/10598405080240020301
- Kunst, M., Bogaerts, S., & Winkel, F. W. (2010). Domestic violence and mental health in a Dutch community sample: The adverse role of loneliness. *Journal of Community & Applied Social Psychology*, 20(5), 419-425. doi: 10.1002/casp.1040
- Ladd, G. W., & Ettekal, I. (2013). Peer-related loneliness across early to late adolescence: Normative trends, intra-individual trajectories, and links with depressive symptoms. *Journal of Adolescence*, 36(6), 1269-1282. doi: 10.1016/j.adolescence.2013.05.004
- Lasgaard, M., Goossens, L., & Elklit, A. (2011). Loneliness, depressive symptomatology, and suicide ideation in adolescence: Cross-sectional and longitudinal analyses. *Journal of Abnormal Child Psychology*, 39(1), 137-150. doi: 10.1007/s10802-010-9442-x
- Laursen, B., & Hartl, A. C. (2013). Understanding loneliness during adolescence: Developmental changes that increase the risk of perceived social isolation. *Journal of Adolescence*, 36(6), 1261-1268. doi: 10.1016/j.adolescence.2013.06.003
- Leung, G. T. Y., De Jong-Gierveld, J., & Lam, L. C. W. (2008). Validation of the Chinese translation of the

- 6-item De Jong-Gierveld Loneliness Scale in elderly Chinese. *International Psychogeriatrics*, 20(6), 1262-1272. doi: 10.1017/S1041610208007552
- Levine, M. P. (2012). Loneliness and eating disorders. *The Journal of Psychology*, 146(1-2), 243-257. doi: 10.1080/00223980.2011.606435
- Levitt, M. J., Guacci-Franco, N., & Levitt, J. L. (1993). Convoys of social support in childhood and early adolescence: Structure and function. *Developmental Psychology*, 29, 811-818. doi:10.1037/0012-1649.29.5.811
- Lorh, S. (2000). *Muestreo: Diseño y análisis*. México, DF: Internacional Thomson Editores.
- Madsen, K. R., Damsgaard, M. T., Rubin, M., Jervelund, S. S., Lasgaard, M., Walsh, S. ... & Holstein, B. E. (2016). Loneliness and ethnic composition of the school class: A nationally random sample of adolescents. *Journal of Youth and Adolescence*, 45(7), 1350-1365. doi: 10.1007/s10964-016-0432-3
- Mahon, N. E., Yarcheski, A., Yarcheski, T. J., Cannella, B. L., & Hanks, M. M. (2006). A meta-analytic study of predictors for loneliness during adolescence. *Nursing Research*, 55(5), 308-315.
- Marcoen, A., Goossens, L., & Caes, P. (1987). Loneliness in pre-through late adolescence: Exploring the contributions of a multidimensional approach. *Journal of Youth and Adolescence*, 16(6), 561-577.
- Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, 57(3), 519-530. doi: 10.1093/biomet/57.3.519
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Erlbaum.
- Mcwhirter, B. T., Besett-Alesch, T. M., Horibata, J., & Gat, I. (2002). Loneliness in high risk adolescents: The role of coping, self-esteem, and empathy. *Journal of Youth Studies*, 5(1), 69-84. doi: 10.1080/13676260120111779
- Meltzer, H., Bebbington, P., Dennis, M. S., Jenkins, R., McManus, S., & Brugha, T. S. (2013). Feelings of loneliness among adults with mental disorder. *Social Psychiatry and Psychiatric Epidemiology*, 48(1), 5-13. doi: 10.1007/s00127-012-0515-8
- Miller, G. (2011). Why loneliness is hazardous to your health. *Science*, 331(6014), 138-140.
- Muñiz, J. (1994). *Teoría clásica de los test*. Madrid: Pirámide.
- Mushtaq, R., Shoib, S., Shah, T., & Mushtaq, S. (2014). Relationship between loneliness, psychiatric disorders and physical health? A review on the psychological aspects of loneliness. *Journal of Clinical and Diagnostic Research*, 8(9), WE01-WE04. doi: 10.7860/JCDR/2014/10077.4828
- Muthén, B., Kaplan, D., & Hollis, M. (1987). On structural equation modeling with data that are not missing completely at random. *Psychometrika*, 52(3), 431-462.
- Page, R. M., Dennis, M., Lindsay, G. B., & Merrill, R. M. (2011). Psychosocial distress and substance use among adolescents in four countries: Philippines, China, Chile, and Namibia. *Youth & Society*, 43(3), 900-930. doi: 10.1177/0044118X10368932
- Parlee, M. B. (1979). The friendship bond. *Psychology Today*, 13, 43-54.
- Perlman, D., & Peplau, L. A. (1981). Toward a social psychology of loneliness. En S. Duck & R. Gilmour (Eds.), *Personal Relationships 3: Personal Relationships in Disorder* (pp. 31-56). New York: Academic Press.
- Pontes, H. M., Griffiths, M. D., & Patrão, I. M. (2014). Internet addiction and loneliness among children and adolescents in the education setting: An empirical pilot study. *Aloma: Revista de Psicologia, Ciències de L'educació i de L'esport Blanquerna*, 32(1), 91-98.
- Qualter, P., Brown, S. L., Rotenberg, K. J., Vanhalst, J., Harris, R. A., Goossens, L. ... & Munn, P. (2013). Trajectories of loneliness during childhood and adolescence: Predictors and health outcomes. *Journal of Adolescence*, 36(6), 1283-1293. doi: 10.1016/j.adolescence.2013.01.005
- Qualter, P., Vanhalst, J., Harris, R., Van Roekel, E., Lodder, G., Bangee, M. ... & Verhagen, M. (2015). Loneliness across the life span. *Perspectives*

- on *Psychological Science*, 10(2), 250-264. doi: 10.1177/1745691615568999
- R Development Core Team (2007). R: A language and environment for statistical computing. [software de cómputo] Viena, Austria: R Foundation for Statistical Computing Disponible en <http://cran.r-project.org/>
- Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment*, 92(6), 544-559. doi: 10.1080/00223891.2010.496477
- Reise, S. P., Scheines, R., Widaman, K. F., & Haviland, M. G. (2013). Multidimensionality and structural coefficient bias in structural equation modeling: A bifactor perspective. *Educational and Psychological Measurement*, 73, 5-26. doi: 10.1177/0013164412449831
- Rios, J., & Wells, C. (2014). Validity evidence based on internal structure. *Psicothema*, 26(1), 108-116. doi: 10.7334/psicothema2013.260
- Rodríguez, A., Reise, S. P., & Haviland, M. G. (2016). Applying bifactor statistical indices in the evaluation of psychological measures. *Journal of Personality Assessment*, 98(3), 223-237. doi: 10.1080/00223891.2015.1089249
- Rodríguez, M. N., & Ruiz, M. A. (2008). Atenuación de la asimetría y de la curtosis de las puntuaciones observadas mediante transformaciones de variables: Incidencia sobre la estructura factorial. *Psicológica: Revista de Metodología y Psicología Experimental*, 29(2), 205-227.
- Rokach, A. (2012). Loneliness updated: An introduction. *The Journal of Psychology*, 146(1-2), 1-6. doi: 10.1080/00223980.2012.629501
- Rubio, R., & Aleixandre, M. (2001). Un estudio sobre la soledad en las personas mayores: Entre estar solo y sentirse solo. *Revista Multidisciplinar de Gerontología*, 11, 23-28.
- Satorra, P., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, 66(4), 507-514.
- Schinka, K. C., Van Dulmen, M. H. M., Bossarte, R., & Swahn, M. (2012). Association between loneliness and suicidality during middle childhood and adolescence: Longitudinal effects and the role of demographic characteristics. *The Journal of Psychology: Interdisciplinary and Applied*, 146, 105-118. doi: 10.1080/00223980.2011.584084
- Schultz, N. R., & Moore, D. (1988). Loneliness: Differences across three age levels. *Journal of Social and Personal Relationships*, 5, 275-284.
- Shiovitz-Ezra, S., & Ayalon, L. (2012). Use of direct versus indirect approaches to measure loneliness in later life. *Research on Aging*, 34(5), 572-591. doi: 10.1177/0164027511423258
- Stucky, B. D., Edelen, M. O., Vaughan, C. A., Tucker, J. S., & Butler, J. (2014). The psychometric development and initial validation of the DCI-A Short Form for adolescent therapeutic community treatment process. *Journal of Substance Abuse Treatment*, 46, 516-521
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5^a ed.). Boston: Pearson/Allyn & Bacon.
- Tiwari, S. C. (2013). Loneliness: A disease? *Indian Journal of Psychiatry*, 55(4), 320-322. doi: 10.4103/0019-5545.120536
- Tomás, J. M., Pinazo-Hernandis, S., Donio-Bellegarde, M., & Hontangas, P. M. (2017). Validity of the De Jong-Gierveld Loneliness Scale in Spanish older population: Competitive structural models and item response theory. *European Journal of Ageing*, 1-9. doi: 10.1007/s10433-017-0417-4
- Van Dulmen, M. H., & Goossens, L. (2013). Loneliness trajectories. *Journal of Adolescence*. 36(6), 1247-1249. doi: 10.1016/j.adolescence.2013.08.001
- Van Roekel, E., Scholte, R. H., Engels, R. C., Goossens, L., & Verhagen, M. (2015). Loneliness in the daily lives of adolescents: An experience sampling study examining the effects of social contexts. *The Journal of Early Adolescence*, 35(7), 905-930. doi: 10.1177/0272431614547049
- Van Tilburg, T., Havens, B., & De Jong-Gierveld, J. (2004). Loneliness among older adults in the Netherlands,

- Italy, and Canada: A multifaceted comparison. *Canadian Journal on Aging*, 23(2), 169-180. doi: 10.1353/cja.2004.0026
- VanderWeele, T. J., Hawkey, L. C., & Cacioppo, J. T. (2012). On the reciprocal association between loneliness and subjective well-being. *American Journal of Epidemiology*, 176(9), 777-784. doi: 10.1093/aje/kws173
- Vanhalst, J., Klimstra, T. A., Luyckx, K., Scholte, R. H., Engels, R. C., & Goossens, L. (2012). The interplay of loneliness and depressive symptoms across adolescence: Exploring the role of personality traits. *Journal of Youth and Adolescence*, 41(6), 776-787. doi 10.1007/s10964-011-9726-7
- Vanhalst, J., Luyckx, K., & Goossens, L. (2014). Experiencing loneliness in adolescence: A matter of individual characteristics, negative peer experiences, or both? *Social Development*, 23(1), 100-118. doi: 10.1111/sode.12019
- Ventura-León, J. L. (2017). ¿Existen los instrumentos válidos? Un debate necesario. *Gaceta Sanitaria*, 31(1), 71-71.
- Ventura-León, J. L., & Caycho, T. (2017). Validez y fiabilidad de la Escala de Soledad de De Jong-Gierveld en jóvenes y adultos peruanos. *PSIENCIA. Revista Latinoamericana de Ciencia Psicológica*, 9. doi: 10.5872/psiencia/9.1.41
- Ventura-León, J. L., & Caycho-Rodríguez, T. (2017). El coeficiente omega: Un método alternativo para la estimación de la confiabilidad. *Revista Latinoamericana de Ciencias Sociales, Niñez y Juventud*. 15(1), 625-627.
- Victor, C. R., & Yang, K. (2012). The prevalence of loneliness among adults: A case study of the United Kingdom. *The Journal of Psychology*, 146(1-2), 85-104. doi: 10.1080/00223980.2011.613875
- Weare, K. (2000). *Promoting Mental: A whole school approach*. New York: Routledge.
- Weeks, M. S., & Asher, S. R. (2012). Loneliness in childhood: Toward the next generation of assessment and research. *Advances in Child Development and Behavior*, 42, 1-39. doi: 10.1016/B978-0-12-394388-0.00001-0
- Whitehouse, A. J., Durkin, K., Jaquet, E., & Ziatas, K. (2009). Friendship, loneliness and depression in adolescents with Asperger's Syndrome. *Journal of Adolescence*, 32(2), 309-322. doi: 10.1016/j.adolescence.2008.03.004
- Woodhouse, S. S., Dykas, M. J., & Cassidy, J. (2012). Loneliness and peer relations in adolescence. *Social Development*, 21(2), 273-293. doi: 10.1111/j.1467-9507.2011.00611.x
- Zammuner, V. L. (2008). Italians' social and emotional loneliness: The results of five studies. *International Journal of Human and Social Sciences*, 3(2), 108-120.
- Zinbarg, R. E., Barlow, D. H., & Brown, T. A. (1997). Hierarchical structure and general factor saturation of the Anxiety Sensitivity Index: Evidence and implications. *d(3)*, 277-284. doi: 10.1037/1040-3590.9.3.277

Inhibición de Borrado en Adultos Mayores: Aportes para la Validación de una Tarea

Deletion Inhibition in Older Adults: Contributions for the Validation of a Task

Ana Comesaña *¹, Florencia Stelzer¹, Isabel Introzzi¹

1- Instituto de Psicología Básica, Aplicada y Tecnología (IPSIBAT) - Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)- Universidad Nacional de Mar del Plata (UNMdP), Argentina.

Recibido: 09/08/2017 **Revisado:** 06/10/2017 **Aceptado:** 13/10/2017

Introducción
Materiales y métodos
Procedimiento
Resultados
Discusión
Referencias

Resumen

La inhibición de borrado constituye un tipo inhibitorio que permite eliminar la información irrelevante durante la ejecución de una tarea en curso. Los instrumentos o pruebas para evaluarla son escasos. El objetivo de este trabajo es obtener evidencia de validez de constructo utilizando una tarea informatizada de la TAC (Tareas de Autorregulación Cognitiva) para medir inhibición de borrado. Para este estudio, se utilizó una muestra de 53 participantes (32 adultos jóvenes y 21 adultos mayores). Los análisis realizados indicaron que la tarea cumple con los supuestos de estructura interna de la tarea y los cambios maduracionales esperados para este proceso. La tarea resulta un instrumento apropiado para la medición de la inhibición de borrado en sujetos adultos jóvenes y mayores.

Palabras clave: *funciones ejecutivas, inhibición de borrado, tarea informatizada, adultos mayores, validación de constructo*

Abstract

Deletion inhibition constitutes an inhibitory kind which allows for the elimination of irrelevant information during the execution of a task. The instruments or tests for evaluating it are scarce. The aim of this work is to obtain evidence of construct validity using a computerized task from the TAC (cognitive auto-regulating task) to measure deletion inhibition. Fifty three adults (32 young adults and 21 older adults) participated in this study. The analyses carried out indicate that the task fulfills the assumptions of internal structure and the maturational changes expected for this process. The task is an appropriate instrument for measuring deletion inhibition in young and older adults.

Keywords: *executive functions, deletion inhibition, computerized task, older adults, construct validity*

*Correspondencia a: Dra. Ana Comesaña a: Matheu 3074 (7600). 54 9 223 560-9511 Mar del Plata, Argentina. acomesan@mdp.edu.ar

Cómo citar este artículo: Comesaña, A., Stelzer, F., & Introzzi, I. (2017). Inhibición de borrado en adultos mayores: Aportes para la validación de una tarea. *Revista Evaluar*, 17(2), 143-153. Recuperado de <https://revistas.unc.edu.ar/index.php/revaluar>

Agradecimientos: Se agradece la colaboración de todos los participantes de esta investigación. Este trabajo se enmarcó en un proyecto financiado por la Secretaría de Ciencia y Tecnología y Coordinación de la Universidad Nacional de Mar del Plata ("Desarrollo de la regulación de la cognición, el comportamiento y la emoción durante el ciclo vital") y fue desarrollado durante los años 2015-2016. También se debe agradecer al CONICET, institución a la cual pertenecen las autoras como investigadoras.

Las autoras manifiestan no tener conflicto de intereses.

Introducción

La *inhibición* contribuye de manera significativa a nuestra vida cotidiana, a la autorregulación y a la adaptación al medio ambiente a través del control de nuestros pensamientos, comportamientos y emociones, nos permite adaptarnos. Si bien no existe una única definición de este proceso, la mayoría de los modelos actuales concuerdan con que su función principal consiste en limitar el ingreso de información irrelevante a la conciencia o foco atencional, lo que permite mantener activa la información relevante para el logro de objetivos (Diamond, 2013; Introzzi, Canet-Juric, Aydmune, & Stelzer, 2016). Por ejemplo, luego de una fuerte discusión es probable que los recuerdos y pensamientos vinculados con ella se presenten de manera intrusiva en nuestra conciencia, compitiendo por los recursos atencionales e interfiriendo con la tarea que estemos realizando (p. ej., una actividad laboral). En esta situación, resulta esencial la habilidad para suprimir o eliminar del foco atencional tales representaciones, dado que estas afectan negativamente nuestro desempeño.

Aunque actualmente existe un acuerdo sobre la importancia de la inhibición para la adaptación y el adecuado funcionamiento de otras funciones cognitivas, se observan divergencias en cuanto a su definición conceptual y operativa. Se distinguen dos enfoques principales: uno que entiende a la inhibición como una función global y unitaria (Cohen, Dunbar, & McClelland, 1990; Morton & Munakata, 2002) y otro que fragmenta la inhibición en un conjunto de procesos con características operativas distintivas: la *inhibición de acceso*, que se ocupa de controlar el acceso de la información al foco atencional, la *inhibición de borrado*, que tiene como objetivo borrar del foco atencional y de la memoria de trabajo la información irrelevante para una tarea en curso

y, finalmente, la *inhibición de restricción*, que se encarga de suprimir o restringir comportamientos prepotentes pero inadecuados para los objetivos actuales. De todas ellas, la inhibición de borrado, también llamada cognitiva, es la que ha recibido la menor atención.

Dentro de la línea de los enfoques fragmentados o multidimensionales (Miyake & Friedman, 2012), uno de los más representativos es la “Teoría o Hipótesis de la Ineficiencia Inhibitoria” propuesta por Hasher y Zacks (1988) y revisada en distintas oportunidades (Hasher, Zacks, & May, 1999; Healey, Ngo, & Hasher, 2014; Lustig, Hasher, & Zacks, 2008). Según esta teoría, la causa subyacente de la disminución en el funcionamiento cognitivo en los adultos mayores es un déficit inhibitorio. Algunos autores sostienen que la afectación en los adultos mayores varía entre los distintos subtipos de inhibición (Vadaga, Blair, & Li, 2015). En relación con la inhibición de borrado durante el envejecimiento, la disminución de su eficiencia se manifiesta en una mayor dificultad para borrar la información irrelevante del foco atencional o para disminuir su activación desde la memoria de largo plazo en comparación con otros grupos etarios (Ortega, Gómez-Arizaga, Román, & Bajo, 2012). Es decir que la información que previamente resultó relevante para el logro de algún objetivo o alguna tarea actualmente tiene que ser olvidada o borrada.

La relevancia de la inhibición para explicar el declive cognitivo asociado con el envejecimiento impulsó el desarrollo de paradigmas destinados a su evaluación. Uno de los instrumentos de evaluación más utilizados es la prueba diseñada según el paradigma del *olvido dirigido con reconocimiento* propuesta por Oberauer (2001; 2005) que es una adaptación del paradigma original de Stenberg (1969) retomada por otros autores (Aslan & Bäuml, 2012; Sahakyan, Delany, & Goodmon, 2008; Seigo, Goldin, & Gottlob,

2006). Para este tipo de pruebas los sujetos deben memorizar dos listas de palabras, luego aparece una señal de olvido que indica cuál de las listas deben retener en su memoria y finalmente se efectúa una prueba de reconocimiento. En la prueba de reconocimiento los sujetos deben identificar la pertenencia o no de una serie de ítems de prueba a los de una lista que debían retener en la memoria. Los ítems de prueba pueden ser relevantes (PR; pertenecían a la lista de estímulos que debían recordar); irrelevantes (PI; pertenecían a la lista de estímulos que debían olvidar) o nuevos (PN; no se hallaban en ninguna de las dos listas). Distintos estudios con este paradigma indicaron mayores tiempos medios de respuesta (TR) para los PI en comparación con los PN. Asimismo, reportaron un menor porcentaje de respuestas correctas frente a los PI respecto de los PN (precisión; Joorman & Gotlib, 2008; Oberauer, 2001). Los PI constituyen información que se torna irrelevante y debe borrarse de la memoria, por lo que las diferencias observadas en los TR y la precisión entre PI y PN se definen como índices para medir la eficiencia operativa de la inhibición de borrado.

Si bien el paradigma desarrollado por Oberauer (2001) constituye un instrumento válido de medición de la capacidad de inhibición de borrado, su utilización en participantes de distinta cultura exige adaptaciones, dado que los estímulos de la prueba original eran verbales. Diferentes estudios han señalado que la extensión fonológica, la familiaridad y frecuencia de uso y la valencia emocional afectan la retención de los conceptos y el nivel de interferencia en la memoria de trabajo (Baddeley, 2012; Cowan et al., 2005). Además, la familiaridad y la frecuencia de uso de ciertas palabras varían para diferentes grupos etarios y ambientes socioculturales (Vivas, Vivas, Comesaña, García-Coni, & Vorano, 2016; Yao, Wu, Zhang, & Wang, 2016). Según nuestros conocimientos, en Latinoamérica no se cuenta con instrumentos

válidos y confiables para la medición de la inhibición de borrado que puedan ser utilizadas en diferentes grupos de edad, y adaptables en diferentes contextos socioculturales. Se optó por diseñar un instrumento con estímulos visuales (figuras abstractas no codificables semánticamente), que permita controlar el efecto de las variables psicolingüísticas en el procesamiento de los estímulos.

Por todo lo anterior, el principal objetivo de este trabajo consiste en presentar evidencia preliminar sobre la validez de constructo de una prueba informatizada diseñada para medir la inhibición de borrado. Esta prueba se enmarca en una batería informatizada (Tareas de Autorregulación Cognitiva, TAC) desarrollada por Introzzi y Canet Juric (Introzzi, Andrés, Canet-Juric, & Richard's, 2016) y diseñada especialmente para evaluar distintos procesos cognitivos. Como explica Hogan (2015), los estudios de validación resultan de especial interés ya que proporcionan evidencia empírica sobre el grado en que las puntuaciones obtenidas representan indicios adecuados del rasgo o constructo de interés. En este trabajo se presentan dos tipos de evidencias de validez: la validez de estructura interna y la validez de cambios maduracionales o del desarrollo (ambas contribuyen a la validez del constructo) utilizando la prueba de inhibición de borrado en cuestión siguiendo los criterios mencionados por Hogan (2015).

Materiales y métodos

Participantes

La muestra inicial fue seleccionada de manera intencional no probabilística y dividida en dos grupos etarios: uno integrado por adultos jóvenes ($n = 35$) de 20 a 40 años y el otro por adultos mayores de más 65 años ($n = 34$), de la ciudad de Mar del Plata, Argentina.

Los criterios de inclusión para ambos gru-

pos fueron: ausencia de diagnóstico de enfermedades psiquiátricas y/o neurológicas, focales o degenerativas (depresión mayor, esquizofrenia, ACV, DCL, esclerosis múltiple, Alzheimer, Parkinson, entre otras) y un nivel de educación formal superior a los siete años. Además, para el grupo de adultos mayores fue un requisito obtener 82 puntos o más en el ACE-R (Adenbrook's Cognitive Examination Revisado; Torralva et al., 2011), ya que este es el puntaje de corte señalado en esta adaptación para sospechar deterioro cognitivo; por esta razón fueron excluidas de este grupo ocho personas. También se excluyeron de la muestra final cinco adultos mayores y tres jóvenes que se alejaban más de tres desvíos típicos del rendimiento promedio de sus respectivos grupos.

La muestra final quedó conformada por 32 adultos jóvenes con una $M_{\text{edad}} = 28.75$, $DE = 6.08$ (65.6% mujeres y 34.4% varones) y por 21 adultos mayores $M_{\text{edad}} = 69.43$ $DE = 5.73$ (66.7% mujeres y 33.3% varones).

Instrumentos

ACE-R. Dada la alta prevalencia de los trastornos cognitivos en los adultos mayores de 60 años, se administró el ACE-R como método de screening para discriminar entre el envejecimiento normal y el patológico y utilizar a este último como criterio de exclusión. El ACE (Addenbrooke's Cognitive Examination; Mathuranath, Nestor, Berrios, Rakowicz, & Hodges, 2000) es un test breve que evalúa distintos dominios cognitivos para diferenciar entre envejecimiento normal y patológico y entre demencia tipo Alzheimer y demencia Frontotemporal. El puntaje máximo es de 100 y evalúa: orientación (10), atención (8), memoria (35), fluencia verbal (14), lenguaje (28) y habilidades visuoespaciales (5). La adaptación del ACE-R en español (Torralva et al., 2011)

mostró un nivel de confiabilidad interna alta (alfa de Cronbach = .89). La validez concurrente, determinada por la correlación entre el ACE-R y el CDR (Clinical Dementia Rating; Hughes, Berg, Danzinger, Coben, & Martin, 1982), fue estadísticamente significativa ($p < .001$) y la concordancia entre evaluadores fue alta ($kappa$ de Cohen = .98).

Prueba de inhibición de borrado visual (TAC).

Esta prueba se diseñó con base en el paradigma de Oberauer (2001) de olvido dirigido y reconocimiento. La prueba original, al igual que la presentada aquí, está estructurada en tres fases: En la *fase de aprendizaje* (fase 1) se deben memorizar dos listas de estímulos (figuras abstractas en rojo o azul) distribuidos en filas de uno a tres elementos localizados unos en la mitad superior de la pantalla y otros en la mitad inferior, los de la mitad superior se presentan siempre en color rojo y los de la mitad inferior en azul. En la *fase señal* (fase 2): se le indica al participante sobre qué lista realizará la *fase de reconocimiento* (lista *target*). Si la señal consiste en un rectángulo azul, en la fase de reconocimiento se evaluará el recuerdo de la lista azul y si la señal es un rectángulo rojo, se evaluará el recuerdo de la lista roja. El 50% de los ensayos evalúa la lista azul y el otro 50% la roja y su distribución es aleatoria a lo largo de toda la prueba. Por último, en la *fase de reconocimiento* (fase 3), el rectángulo se mantiene en la pantalla y en el centro aparece una figura en color blanco -ítem de prueba- y los participantes deben decidir si forma parte o no de la lista *target* con la mayor rapidez y precisión posible presionando la tecla "Z" (si considera que estaba en la lista *target*) y la tecla "M" (si no estaba en esa lista). Se distinguen tres tipos de ítems de prueba: el 50% son ítems o *probes* relevantes o pertenecientes a la lista *target* (PR), el 25% son ítems o *probes* intrusos o pertenecientes a la lista irrelevante o por ser borrada (PI) y el 25% restante son ítems o *probes* nuevos (PN) o no contenidos en ninguna de las dos listas.

Estructura de los ensayos. Cada ensayo comienza con un punto de fijación que permanece en pantalla durante 200 milisegundos (ms); luego se presentan las dos listas con las figuras abstractas (ubicadas en la parte superior e inferior de la pantalla). Por cada figura se computan 1300 ms, por lo que el tiempo de visualización es el producto de la multiplicación del número total de figuras por 1300 ms. Luego de un intervalo de 800 ms, aparece un rectángulo (azul o rojo) en el centro de la pantalla durante 200 ms. Por último, el tiempo que separa la señal del ítem de prueba, que aparece en color blanco, es de 300 ms (ver Fig.1). Los principales índices de desempeño de esta prueba son: (a) las diferencias entre el tiempo medio de respuesta (calculado solamente para las respuestas correctas) en el reconocimiento de los PI y el tiempo medio de respuesta en el reconocimiento de los PN (TR medio PI - TR medio PN); (b) la diferencia en el porcentaje de respuestas correctas entre el reconocimiento de los PN y de los PI (respuestas correctas PN - respuestas correctas PI). Cuanto mayor es la diferencia en a y b, menor es la eficiencia del funcionamiento de la inhibición de borrado puesto que para obtener un buen desempeño, el participante debe ser capaz de eliminar de su memoria de trabajo los PI que corresponden a la lista irrelevante tras la señalización.

Cambios sobre la versión original. A la versión original de la prueba se le realizaron algunos cambios para adaptarla a nuestro medio: (a) se reemplazaron los estímulos verbales por estímulos visuales -figuras abstractas- a fin de dificultar el uso de estrategias de procesamiento verbal de los estímulos y (b) se redujo la cantidad de ensayos que integraban la prueba original con el propósito de disminuir el probable efecto de fatiga debido a la extensión (eran 64 ensayos precedidos por 12 de práctica y en la versión actual son 32 ensayos precedidos por 5 de práctica). Los estímulos no verbales fueron seleccionados de entre un total

de 200 figuras abstractas diseñadas especialmente para esta prueba. Estas fueron sometidas a una evaluación de 20 participantes a los que se les solicitó que nombraran la imagen que veían en cada una de ellas. Fueron eliminadas aquellas figuras en las que coincidieron en su denominación dos o más participantes. Finalmente, quedaron 100 figuras que fueron utilizadas como estímulos para el diseño de los 32 ensayos de la tarea. Cada ensayo presenta tres fases sucesivas: aprendizaje, señal y reconocimiento (ya mencionadas en el apartado anterior).

Se muestran ejemplos de los tipos de ítems de prueba y las tres fases de la tarea para medir la inhibición de borrado de la batería informatizada TAC (Introzzi, Andrés, et al., 2016) junto con los tiempos de presentación en la Fig.1 a continuación.

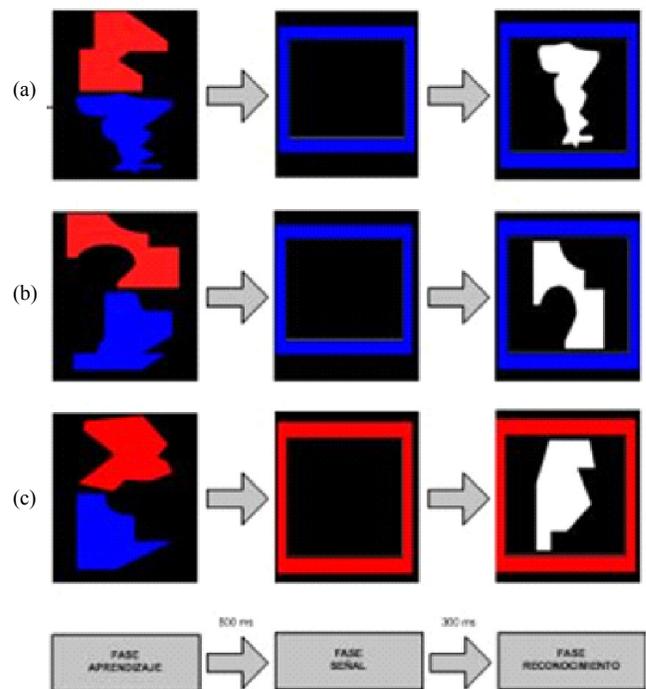


Figura 1

Ejemplo de secuencia de tres ensayos con listas de 1x1 estímulos de la tarea TAC.

Nota. Ejemplos de ensayos con distintos ítems de prueba (a) Relevante, (b) Irrelevante y (c) Nuevo, y los tres tipos de fases de la tarea.

Procedimiento

La participación fue voluntaria. Primero se les presentó a los participantes el consentimiento informado que explicaba los objetivos, las pruebas a ser administradas en esta investigación, el tratamiento y el uso confidencial de los datos conforme con la Declaración de Helsinki (Asociación Médica Mundial, 2013), y en consonancia con los principios éticos y el código de conducta de los psicólogos establecido y reformulado por la APA (2017). También se siguieron los procedimientos indicados por la Ley Nacional Argentina N°25.326 (2000) de protección de datos personales reglamentada por el decreto 1158/2001. La administración se realizó en una entrevista individual con una duración de entre 20 a 30 minutos. Se aplicó la prueba de borrado visual de la TAC (Introzzi, Andrés, et al, 2016), en una computadora HP LAPTOP- RJSENA2U con Windows 10 y una pantalla de 15,6" para ambos grupos. Además, al grupo de adultos mayores se les administró en primer lugar el test ACE-R (Torralva et al., 2011), para evaluar las funciones cognitivas y asegurar que superaran el puntaje de corte establecido para su inclusión en este estudio.

Análisis de los datos

Conforme a lo señalado en la literatura (Oberauer, 2001; Joorman & Gotlib, 2008) se considerarán como criterios de validez interna que (a) el TR medio en el reconocimiento de los PI debe ser superior al TR medio en el reconocimiento de los PN, debido a que los primeros generan una mayor interferencia, y (b) el porcentaje de respuestas correctas debe ser superior en el reconocimiento de los PN respecto al reconocimiento de los PI. Por otra parte, se espera que el grupo de adultos mayores presente una menor eficiencia inhibitoria

que el grupo de jóvenes, lo cual se podrá observar en tanto en mayores TRs como en un menor porcentaje de respuestas correctas para todos los tipos de *probes*. Para realizar los análisis estadísticos se utilizó el paquete SPSS 23.0 de IBM (2014). Para calcular los TR medios se eliminaron los *outliers*, los TR individuales que estuvieran a más de 3 DE de la media del sujeto.

Se exploró la distribución normal de las variables en cada grupo a través de la prueba de Kolmogorov-Smirnov. En los adultos jóvenes todas las variables (TR en PI y en PN; *precisión* en PN y en PI; diferencias en TR entre PI y PN y diferencias en precisión entre PN y PI) tienen un valor $p < .05$ por lo que se asume que su distribución no es normal. En el caso del grupo de adultos mayores, las variables de *precisión* (tanto en PN como en PI), y de *diferencias en precisión* entre PN y PI tienen un valor $p < .05$, no siendo normal su distribución; pero para las variables correspondientes a los TR (tanto en PN como en PI y las diferencias entre ellos) sí se puede asumir una distribución normal ya que ambas tienen un valor $p > .05$. Por lo tanto, se utilizaron pruebas no paramétricas (prueba de suma de rangos de Wilcoxon) para el análisis de la mayoría de las variables salvo para las referidas a los TR en el grupo de adultos mayores, en el que se utilizó una prueba t para muestras relacionadas. Para comparar el rendimiento entre el grupo de adultos mayores respecto al de jóvenes en (a) el TR medio en PI - TR medio en PN y (b) el porcentaje de respuestas correctas en PN - respuestas correctas en PI, se utilizó el Test de U de Mann Whitney para muestras independientes.

Resultados

En primer lugar se realizó el análisis descriptivo de los principales índices de la tarea de borrado, que se muestran en la Tabla 1.

Tabla 1

Porcentaje de respuestas correctas y TR medios en PN y PI para los grupos de jóvenes y mayores.

	Grupo	Media	Desviación estándar	Mediana
TR en PI	Jóvenes	1820.64	833.66	1569.93
	Mayores	3950.93	1901.28	3871.60
TR en PN	Jóvenes	1454.70	728.76	1150.51
	Mayores	3022.34	1361.63	2946.17
Precisión en PN	Jóvenes	96.84	7.88	100.00
	Mayores	82.43	18.60	83.00
Precisión en PI	Jóvenes	81.72	18.12	83.00
	Mayores	68.24	21.67	67.00
Diferencia TR PI- PN	Jóvenes	365.94	428.20	219.00
	Mayores	928.68	736.25	737.40
Diferencia precisión PN- PI	Jóvenes	15.13	16.57	17.00
	Mayores	14.19	18.43	11.43

En la Tabla 1 se puede observar que dentro de cada grupo etario el porcentaje de respuestas correctas (precisión) fue superior en PN respecto de PI. Asimismo, el TR en PI fue superior respecto de la condición PN dentro de cada grupo etario.

A fin de comprobar si las diferencias son estadísticamente significativas y si se cumplió con los criterios de validez interna de la tarea, se efectuaron pruebas *t* y pruebas de suma de rangos de Wilcoxon para muestras relacionadas según la distribución de los datos fuera normal o no. Para ambos grupos, se pudo establecer que las diferencias eran significativas según los criterios de validez interna establecidos y mencionados al inicio del apartado anterior (grupo jóvenes: diferencias en TR PN-PI: $z = 4.93$ $p < .001$ diferencias porcentaje de RC PN-PI: $z = 3.89$, $p < .001$; grupo mayores: diferencias TR PN-PI: $t = 5.77$, $gl: 20$, $p < .001$; diferencias porcentaje de RC PN-PI: $z = 2.82$, $p < .005$).

Verificación del supuesto de validez a través de cambios maduracionales en el desarrollo

Como ya se mencionó, se espera que los adultos mayores presenten una menor eficiencia inhibitoria que los jóvenes. Esto se reflejaría en el hallazgo de una disminución significativa en el grupo de adultos mayores respecto al de jóvenes en (a) el TR medio en PI - TR medio en PN y (b) el porcentaje de respuestas correctas en PN - respuestas correctas en PI. Para contrastar dicho supuesto se utilizó el Test de U de Mann Whitney para muestras independientes.

El análisis de las diferencias en las distribuciones en precisión en PN y en PI entre los dos grupos indicó que estas se deben a circunstancias ajenas a los cambios maduracionales ($z = -0.693$ $p > .05$). Por otro lado, el análisis de las diferencias entre ambos grupos para TR en PI y en PN indicó que éstas no se deben a factores externos a las variables estudiadas ($z = -2.85$; $p < .05$), siendo mayores los tiempos en el grupo de adultos mayores que en el de los jóvenes.

Discusión

El objetivo de este trabajo fue presentar evidencia sobre la validez de constructo de una prueba informatizada desarrollada según el modelo de Oberauer (2001), para evaluar la inhibición de borrado en adultos. La inhibición de borrado permite suprimir la información irrelevante que se mantiene activada en el foco atencional y que interfiere con el objetivo de la tarea en curso, incrementando así la eficacia del procesamiento. La teoría postula que los ítems que se vuelven irrelevantes en la fase de señal de la tarea y deben ser borrados generan interferencia en el reconocimiento de los ítems relevantes en la fase de prueba de la tarea, lo cual se evidencia en mayores TR y en un menor porcentaje de respuestas correctas en los ítems irrelevantes que ante ítems nuevos. Los resultados obtenidos se corresponden con dicha explicación, lo cual indica que la tarea de borrado visual constituye un procedimiento válido para la medición de la inhibición de borrado en adultos.

Con respecto al criterio de validez por cambios maduracionales o en el desarrollo, se observó que las diferencias en los TR entre las condiciones PI y PN son estadísticamente significativas entre el grupo de los adultos jóvenes y el grupo de adultos mayores, en ambas condiciones los jóvenes obtuvieron menores TR. Estos resultados indicarían que los adultos mayores tienen una mayor dificultad para borrar la información irrelevante de la memoria de trabajo, lo que se corresponde con las predicciones de la Teoría de la Ineficiencia Inhibitoria de Hasher y Zacks (1988), que sostiene que los adultos mayores en general presentan una menor eficacia en la inhibición activa de la información que los adultos jóvenes. No obstante, de manera coherente o compatible con otros estudios, los resultados obtenidos mostraron una ausencia de diferencias significativas para el índice basado en los porcentajes de respuestas

correctas (Joorman & Gotlib, 2008; Oberauer, 2001), lo que indicaría que la precisión no es un indicador adecuado para la estimación de cambios cognitivos asociados con el envejecimiento.

Por otra parte, cabe destacar que en nuestro trabajo las tasas de respuesta correctas fueron sensiblemente inferiores a las reportadas en otros estudios (Loosli, Rahm, Unterrainer, Weiller, & Kaller, 2014; Joorman & Gotlib, 2008) posiblemente debido al carácter abstracto de los estímulos utilizados, pero los resultados presentaron, sin embargo, el mismo sentido que los estudios previos mencionados. Esta afirmación se ve respaldada por el hecho de que tanto en el grupo de adultos mayores como en el de jóvenes la tarea de inhibición de borrado mostró un adecuado ajuste a los criterios de validez interna.

En síntesis, el estudio aporta evidencia empírica de validez en relación con la tarea de borrado visual. Sin embargo, es pertinente señalar algunas limitaciones. En primer lugar, la muestra participante fue seleccionada según disponibilidad y su tamaño es pequeño. Serán necesarios posteriores estudios que incrementen el tamaño de la muestra y la representatividad de la misma a través de su selección aleatoria.

Asimismo, el contraste de otras formas de validez tales como la convergente y la divergente, la predictiva y la capacidad de la misma para detectar cambios maduracionales en otros grupos etarios contribuirán a una validación más amplia de la tarea.

La teoría predice que la inhibición de borrado experimenta una mejoría desde la infancia hasta la adultez (Davidson, Amso, Anderson, & Diamond, 2006; Introzzi, Canet Juric et. al, 2016), por lo que sería pertinente contrastar si la tarea desarrollada es válida para su aplicación en niños y adolescentes. Otros integrantes del mismo grupo de investigación han realizado estudios para validar esta y otras pruebas de la batería informa-

tizada TAC en poblaciones infantiles, y de adultos jóvenes (Introzzi, Andrés, et al., 2016; Richard's, Introzzi, Zamora & Vernucci, 2017)

Tal como fue mencionado en la introducción, la inhibición constituye un proceso cognitivo clave para la adaptación de los sujetos al ambiente. Si bien la literatura actual distingue diferentes tipos inhibitorios, en Argentina no habían sido desarrolladas hasta el momento tareas que permitan la medición de la inhibición de borrado. El disponer de instrumentos válidos para la medición de los diferentes tipos inhibitorios permite una mejor comprensión de la relación de estos procesos con otros procesos cognitivos, así como con variables académicas y la semiología de diferentes trastornos mentales. Con este estudio aspiramos a contribuir a tal meta al presentar una tarea con evidencias de validez para la medición de uno de los componentes de la inhibición, la de borrado.

Referencias

- American Psychological Association. (2017). *Ethical principles of psychologists and code of conduct*. Recuperado de <http://www.apa.org/ethics/code/ethics-code-2017.pdf>
- Aslan, A., & Bäuml, K. H. T. (2012). Retrieval-induced forgetting in old and very old age. *Psychology and Aging, 27*(4), 1027-1032. doi: 10.1037/a0028379
- Asociación Médica Mundial (2013). *Declaración de Helsinki. Principios éticos para las investigaciones con seres humanos*. Recuperado de <https://www.wma.net/es/policias-post/declaracion-de-helsinki-de-la-amm-principios-eticos-para-las-investigaciones-medicas-en-seres-humanos>
- Baddeley, A. (2012). Working memory: Theories, models and controversies. *Annual Review of Psychology, 63*(1), 1-29. doi: 10.1146/annurev-psych-120710-100422
- Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed account of the Stroop effect. *Psychological Review, 97*(3), 332-361. doi: 10.1037/0033-295X.97.3.332
- Cowan, N., Elliot, E. M., Scott-Saults, J. S., Morey, C., Mattox, S., Hismajatullina, A., & Conway, A. (2005). On the capacity of attention: Its estimation and role in working memory and cognitive aptitudes. *Cognitive Psychology, 51*(1), 42-100. doi: 10.1016/j.cogpsych.2004.12.001
- Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia, 44*(11), 2037-2078. doi: 10.1016/j.neuropsychologia.2006.02.006
- Diamond, A. (2013). Executive functions. *Annual Review of Psychology, 64*(1), 135-168. doi: 10.1146/annurev-psych-113011-143750
- Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. *The Psychology of Learning and Motivation, 22*, 193-225. doi: 10.1016/s0079-7421(08)60041-9
- Hasher, L., Zacks, R. T., & May, C. P. (1999). Inhibitory control, circadian arousal, and age. En D. Gopher & A. Koriat (Eds.), *Attention & Performance, XVII, Cognitive Regulation of Performance: Interaction of Theory and Application* (pp. 653-675). Cambridge, MA: MIT Press. Recuperado de http://www.psych.utoronto.ca/users/haserlab/abstracts/haser_zacks_may99.htm
- Healey, M. K., Ngo, K. W. J., & Hasher, L. (2014). Below-baseline suppression of competitors during interference resolution by younger but not older adults. *Psychological Science, 25*, 145-151. doi: 10.1177/0956797613501169
- Hogan T. R. (2015). Validez. En T. R. Hogan (Ed.), *Pruebas psicológicas: Una introducción práctica* (2ª ed., pp. 199-258). México, DF: Manual Moderno.
- Hughes, C. P., Berg, L., Danzinger, W. L., Coben, L. A., & Martin, R. L. (1982). A new clinical scale for staging of dementia. *The British Journal of Psychiatry,*

- 140(6), 566-572. doi: 10.1192/bjp.140.6.566
- IBM Corp (2015). IBM SPSS Statistics for Windows, versión 23.0 [software de cómputo]. Armonk, N.Y. USA.
- Introzzi, I., Andrés, M. L., Canet-Juric, L., & Richard's, M. M. (2016). The relationship between the rumination style and perceptual, cognitive and behavioral inhibition. *Psychology & Neuroscience*, 9(4), 444-456. doi: 10.1037/pne0000068
- Introzzi, I. M., Canet-Juric, L., Aydmune, Y., & Stelzer, F. (2016). Perspectivas teóricas y evidencia empírica sobre la inhibición. *Revista Colombiana de Psicología*, 25(2), 351-368. doi: 10.15446/rcp.v25n2.52011
- Joorman, J., & Gotlib, I. H. (2008). Updating the contents of working memory in depression: Interference from irrelevant negative material. *Journal of Abnormal Psychology*, 117(1), 182-92. doi: 10.1037/0021-843X.117.1.182
- Ley N°25. 326. Protección de datos personales. Dirección Nacional de Protección de datos personales, Ministerio de Justicia, Seguridad y Derechos humanos, 29 de Noviembre de 2001.
- Loosli, S. V., Rahm, B., Unterrainer, J. M., Weiller, C., & Kaller, C. P. (2014). Developmental change in proactive interference across the life span: Evidence from two working memory tasks. *developmental Psychology*, 50(4), 1060. doi: 10.1037/a0035231
- Lustig, C., Hasher, L., & Zacks, R. T. (2008). Inhibitory deficit theory: Recent development in a "new view". En D. S. Gorfein & C. M. MacLeod (Eds.), *The place of inhibition in cognition* (pp. 145-162). Washington, DC: American Psychological Association. doi: 10.1037/11587-008
- Mathuranath, P. S., Nestor, P. J., Berrios, G. E., Rakowicz, W., & Hodges, J. R. (2000). A brief cognitive test battery to differentiate Alzheimer's disease and frontotemporal dementia. *Neurology*, 55(11), 1613-1620. doi: 10.1212/01.wnl.0000434309.85312.19
- Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science*, 21(1), 8-14. doi: 10.1177/0963721411429458
- Morton, J. B., & Munakata, Y. (2002). Active versus latent representations: A neural network model of perseveration, dissociation, and decalage. *Developmental Psychobiology*, 40(3), 255-265. doi: 10.1002/dev.10033
- Oberauer, K. (2001). Removing irrelevant information from working memory: A cognitive aging study with the modified Sternberg task. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 27(4), 948-957. doi: 10.1037/0278-7393.27.4.948
- Oberauer, K. (2005). Control of the contents of working memory –A comparison of two paradigms and two age groups. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 31(4), 714-728. doi: 10.1037/0278-7393.31.4.714
- Ortega, A., Gómez-Ariza, C. J., Román, P., & Bajo, M. T. (2012). Memory inhibition, aging and the executive deficit hypothesis. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 38(1), 178-186. doi: 10.1037/a0024510
- Richard's, M., Introzzi, I., Zamora, E. & Vernucci, S. (2017). Analysis of Internal and External Validity Criteria for a Computerized Visual Search Task. A pilot study. *Applied Neuropsychology: Child*, 6(2) 110-119 (print), 2162-2973 (Online). doi: 10.1080/21622965.2015.1083433
- Sahakyan, L., Delaney, P. F., & Waldum, E. R. (2008). Intentional forgetting is easier after two "shots" than one. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(2), 408. doi: 10.1037/0278-7393.34.2.408
- Sego, S. A., Golding, J. M., & Gottlob, L. R. (2006). Directed forgetting in older adults using the item and list methods. *Aging, Neuropsychology, and Cognition*, 13, 95-114. doi: 10.1080/138255890968682
- Sternberg, S. (1969). Memory scanning: Mental processes revealed by reaction-time experiments. *American Scientist*, 57, 421-457. Recuperado de <http://www.psych.upenn.edu/~saul>

- Torrvalva, T., Roca, M., Gleichgerrcht, E., Bonifacio, A., Raimondi, C., & Manes F. (2011). Validación de la versión en español del Addenbrooke's Cognitive Examination-Revisado (ACE-R). *Neurología*, 26(6), 351-356. doi: 10.1016/j.nrl.2010.10.013
- Vadaga, K. K., Blair, M., & Li, K. Z. H. (2015). Are age-related differences uniform across different inhibitory functions? *Journal of Gerontology: Serie B: Psychological Sciences and Social Sciences*, 71(4), 641-649. doi: 10.1093/geronb/gbv002
- Vivas, J., Vivas, L., Comesaña, A., Coni, A. G., & Vrano, A. (2016). Spanish semantic feature production norms for 400 concrete concepts. *Behavior Research Methods*, 49(3), 1095-1106. doi: 10.3758/s13428-016-0777-2
- Yao, Z., Wu, J., Zhang, Y., & Wang, Z. (2016). Norms of valence, arousal, concreteness, familiarity, imageability, and context availability for 1100 chinese words. *Behavior Research Methods*, 49(4), 1374-1385. doi: 10.3758/s13428-016-0793-2
-

Estructura factorial del Teacher Interpersonal Self-Efficacy Scale en docentes dominicanos: ¿uno o tres factores?

Factorial structure of the Teacher Interpersonal Self-Efficacy Scale for Dominican Teachers: one or three factors?

Dinorah García ¹, Consuelo Cerviño ², Irene Fernández ^{3*}, José Manuel Tomás ³

1 - Instituto de Investigación Superior Pedro Poveda, República Dominicana.

2 - Departamento de Psicología Evolutiva y de la Educación, Facultad de Psicología, Universidad de Valencia, España.

3 - Departamento de Metodología de las Ciencias del Comportamiento, Facultad de Psicología, Universidad de Valencia, España.

Introducción
Método
Resultados
Discusión y conclusiones
Referencias

Recibido: 16/08/2017 Revisado: 20/09/2017 Aceptado: 05/10/2017

Resumen

La autoeficacia docente se define como la creencia específica acerca de la externalidad o internalidad del control de las actividades que realiza el profesor en su trabajo. El objetivo del presente estudio es contribuir a entender las propiedades psicométricas de la escala 'Teacher Interpersonal Self-Efficacy Scale' en una muestra de docentes de República Dominicana. Se pusieron a prueba los modelos factoriales propuestos en la literatura mediante análisis factorial confirmatorio: a) un único factor de *autoeficacia*; b) dos factores de autoeficacia percibida: en la *gestión del aula* y en la *obtención de apoyo*; y c) tres factores de autoeficacia percibida: en la *gestión del aula*, en la *obtención de apoyo de los compañeros*, y en la *obtención de apoyo de los superiores*. Los tres modelos mostraron un ajuste que varía entre bueno y excelente. Se retuvo el modelo de tres factores. Se han incluido las implicaciones y limitaciones derivadas de la investigación.

Palabras clave: *autoeficacia docente, profesores dominicanos, estructura factorial*

Abstract

Teacher self-efficacy is defined as the specific belief about the degree to which activities carried out by the subject are either externally or internally controlled. The aim of this research is to contribute to the understanding of the psychometric properties of the Teacher Interpersonal Self-Efficacy Scale in a sample of teachers of the Dominican Republic. All models documented in the literature were tested by means of confirmatory factor analysis: a) a single factor of *self-efficacy*; b) two factors of perceived self-efficacy: in *classroom management* and in *eliciting support*, and c) three factors of perceived self-efficacy: in *classroom management*, in *eliciting support from colleagues*, and in *eliciting support from school principals*. All three models presented a good-to-excellent fit. The three-factor model was retained. Implications and limitations derived from the research have been included.

Keywords: *teacher self-efficacy, Dominican teachers, factorial structure*

*Correspondencia a: I. Fernández. Avd. Blasco Ibáñez, 21, 46010, Valencia, España. ifermar@outlook.com

Cómo citar este artículo: García, D., Cerviño, C., Fernández, I., & Tomás, J. M. (2017). Estructura factorial del Teacher Interpersonal Self-efficacy Scale en docentes dominicanos: ¿uno o tres factores? *Revista Evaluar*, 17(2), 154-164. Recuperado de <https://revistas.unc.edu.ar/index.php/revaluar>

Introducción

El concepto de *autoeficacia* surge a partir de la teoría socio-cognitiva de Bandura y la teoría del aprendizaje social de Rotter (Chacón, 2006). Siguiendo la línea teórica de Bandura, la autoeficacia se relaciona con los juicios que cada sujeto se forma sobre su propia capacidad para realizar una actividad específica. Así, la conducta humana se ve regulada por la auto-influencia, que se compone principalmente de la creencia sobre la eficacia personal (Bandura, 2000). Lo anterior nos permite pensar que, a menos que uno mismo se crea capaz de producir el efecto deseado mediante una acción, tiene poco incentivo para actuar. De esta forma, la autoeficacia percibida posee un rol importante en el incentivo potencial de los resultados esperados. La autoeficacia se relaciona con la conducta a través de procesos cognitivos, motivacionales, afectivos y de selección (Prieto, 2007). Las creencias de autoeficacia pueden influir en los patrones de pensamiento y actuar como un incentivo o un inhibidor de la conducta. Además, la autoeficacia influye en la competencia, de manera que para que un individuo actúe de forma competente necesitará un equilibrio permanente entre su percepción de autoeficacia y los conocimientos y habilidades que requiere para realizar con éxito una acción específica (Velázquez, 2012). Así, la conducta se adquiere y regula mediante el sentimiento personal de autoeficacia (Gonzalo & León, 1999).

Cuando aplicamos estos conceptos al espacio escolar, naturalmente surge el interés por la autoeficacia docente, que comenzó entendiéndose como el grado en el que el docente se siente capaz de incidir en el rendimiento de sus estudiantes (Bergman, McLaughlin, Bass, Pauly, & Zellman, 1977). Sin embargo, más recientemente se ha conceptualizado la autoeficacia docente como la creencia específica acerca de la externalidad o in-

ternalidad del control de las actividades que realiza el sujeto (Fernández, 2012). Según Bandura (1993), el sentimiento de autoeficacia de los docentes para motivar y promover el aprendizaje afecta los tipos de ambiente de aprendizaje que forman, y el éxito académico que sus estudiantes son capaces de alcanzar. Así, el sentimiento de autoeficacia influye directamente en las prácticas pedagógicas que los docentes llevan a cabo (Chacón, 2006). A su vez, la autoeficacia se ve alterada por una serie de elementos, de acuerdo con lo establecido por Prieto (2007), entre los cuales se encuentran: el nivel de rendimiento de los estudiantes, la edad de los estudiantes, el nivel de preparación del profesorado y el tamaño de la clase. La autoeficacia es directamente proporcional a los tres primeros elementos e indirectamente proporcional al cuarto elemento, esto es, al tamaño de la clase. Con todo, se puede concebir a los docentes con un sentimiento de autoeficacia alto como aquellos que se interesan por el aprendizaje de sus estudiantes, experimentan libertad para tomar decisiones y manifiestan satisfacción con su tarea.

Los instrumentos utilizados para medir la autoeficacia docente incluyen: Teacher Locus of Control (TLC; Rose & Meadway, 1981), Responsibility for Student Achievement (RSA; Guskey, 1981), Webb Efficacy Scale (WES; Ashton, Olejnik, Crocker, & McAuliffe, 1982), Teacher Efficacy Scale (TES; Gibson & Dembo, 1984), Science Teaching Efficacy Belief Instrument (STEBI; Riggs & Enochs, 1990), Ashton Vignettes (Ashton, Buhr, & Crocker, 1984) y Teacher Self-Efficacy Scale (TSS; Bandura, 1990). De todos estos, el Teacher Efficacy Scale (Gibson & Dembo, 1984) es uno de los instrumentos más utilizados según diversos autores (Henson, Kogan, & Vacha-Haase, 2001; Brouwers, Tomic, & Stijnen, 2002). En el desarrollo de esta escala, los autores asumieron que

el constructo de autoeficacia comprendía tanto la autoeficacia docente general –la influencia que un profesor cree que los profesores en general tienen sobre los alumnos– como la personal –la influencia que un docente en particular cree poseer sobre sus alumnos. Sin embargo, la sensación de autoeficacia general resulta problemática puesto que, según lo argumentado por [Bandura \(1997\)](#), las creencias de autoeficacia suelen variar para cada individuo según las tareas específicas por las que se pregunte. Con el fin de solventar esta controversia, se desarrolló la escala ‘Teacher Interpersonal Self-Efficacy Scale’ (TISES; [Brouwers & Tomic, 2001](#)).

La escala ‘Teacher Interpersonal Self-Efficacy Scale’ comprende tres actividades interpersonales de los docentes: la gestión del comportamiento del alumnado en clase, la obtención del apoyo docente y la obtención del apoyo de los superiores. Esta escala ha sido solamente validada en Holanda ([Brouwers & Tomic, 2001](#)) y en España ([García-Ros, Fuentes, & Fernández, 2015](#)). La escala original en holandés fue validada por [Brouwers y Tomic \(2001\)](#), empleando una muestra de calibración y otra de validación. Se pusieron a prueba un modelo nulo (base de comparación), un modelo de un factor, otro de dos factores correlacionados y finalmente uno de tres factores correlacionados. En ambas muestras el modelo de tres factores correlacionados, tras la depuración de un ítem, mostró el mejor ajuste, si bien este puede considerarse sólo marginalmente bueno. Por su parte, en la validación de la escala en alumnos españoles ([García-Ros et al., 2015](#)) se pusieron a prueba tres estructuras factoriales, la teórica –o a priori– de tres factores correlacionados, una estructura de dos factores correlacionados y la más parsimoniosa de un solo factor. El ajuste de los modelos no fue plenamente satisfactorio, aunque mediante la estimación de cinco covarianzas entre los errores se consiguió

un ajuste satisfactorio del modelo de tres factores correlacionados.

El objetivo de esta investigación es contribuir a entender las propiedades psicométricas de la escala ‘Teacher Interpersonal Self-Efficacy Scale’, y en concreto la capacidad de generalización de esta a muestras de docentes de otras poblaciones, concretamente a docentes de República Dominicana. Los objetivos específicos son: (a) analizar la estructura factorial mediante técnicas confirmatorias, poniendo a prueba todos los modelos alternativos que se han empleado en la literatura para la escala y (b) estimar la consistencia interna de la escala. Todos estos análisis se realizarán en una muestra probabilística de docentes dominicanos.

Método

Muestra

La muestra consistió en un total de 633 docentes, de los cuales un 49.9% pertenecían al Distrito 5 de La Vega y el 50.1% restante al Distrito 3 de Santo Domingo, en la República Dominicana. El procedimiento de obtención de la muestra fue de carácter aleatorio por conglomerados. Los distritos educativos de Santo Domingo y La Vega fueron los conglomerados, se escogieron de entre estos los 3 y 5 al azar, y de ellos se extrajo la muestra de docentes. La muestra está compuesta por un 79.5% de mujeres y un 20.5% de hombres. En cuanto al estado civil, la muestra presenta un 63.7% de casados, un 16.6% de solteros y el resto, 19.7%, en situaciones alternativas.

Instrumentos

Para el propósito de este estudio psicométrico, se tuvo en cuenta la escala Teacher Inter-

personal Self-Efficacy Scale (Brouwers & Tomic, 2001). Esta escala consta de 24 ítems, medidos mediante una escala Likert de cinco anclajes, desde 1 (*Totalmente en desacuerdo*) a 5 (*Totalmente de acuerdo*). El instrumento mide a priori tres dimensiones de autoeficacia percibida: gestión del aula, obtención de apoyo de los compañeros y obtención de apoyo de los superiores. La primera dimensión, autoeficacia percibida en la *gestión del aula*, está compuesta por los ítems 1, 4, 5, 8, 9, 10, 11, 13, 14, 15, 17, 18, 22 y 24. Un ítem característico de esta dimensión es, por ejemplo: *Si un alumno interrumpe la clase, soy capaz de re-conducirlo rápidamente*. Estos ítems fueron desarrollados por Emmer y Hickman (1991) y posteriormente adaptados a la sub-escala por Brouwers y Tomic (2001). La segunda dimensión, autoeficacia percibida en la *obtención de apoyo de los compañeros*, incluye los ítems 3, 7, 20, 21 y 23. Un ítem propio de este factor es: *Si me enfrento a un problema en el que mis compañeros me pueden ayudar, soy capaz de acercarme a ellos para ello*. La tercera y última dimensión, autoeficacia percibida en la *obtención de apoyo de los superiores*, comprende los ítems 2, 6, 12, 16 y 19. Un ejemplo de estos ítems es: *Cuando es necesario, soy capaz de sacar a colación problemas con los directivos*. Estas dos dimensiones relacionadas con el apoyo fueron desarrolladas por Brouwers (2000).

Como criterios en una red nomológica se han considerado otros constructos medidos en la misma investigación, Loyola Generativity Scale (LGS; McAdams & St Aubin, 1992) y Escala de Motivaciones Psicosociales (Fernández, 1987). La LGS tiene un primer factor que mide la *expresión generativa* y un segundo factor de *carencia generativa*. Los índices alfa de esta escala fueron .89 y .75, respectivamente. Por su parte, la Escala de Motivaciones Psicosociales está formada por tres factores: *motivación intrínseca*, con alfa de

.839, *motivación extrínseca*, con alfa de .79, y *motivación intrínseca para el desarrollo profesional*, con alfa de .67.

Análisis estadísticos

En esta investigación se evaluaron los modelos puestos a prueba en el artículo original de Brouwers y Tomic (2001) mediante análisis factorial confirmatorio (AFC). El Modelo 1 consta de un único factor de autoeficacia. El Modelo 2 está compuesto por dos factores: autoeficacia percibida en la *gestión del aula* y autoeficacia percibida en la *obtención de apoyo*. Y el Modelo 3 cuenta con tres factores: autoeficacia percibida en la *gestión del aula*, autoeficacia percibida en la *obtención de apoyo de los compañeros*, y autoeficacia percibida en la *obtención de apoyo de los superiores*. Los modelos se muestran en la Figura 1. Se calcularon varios índices de ajuste, tal y como recomienda la literatura (Hu & Bentler, 1999; Tanaka, 1993): el estadístico chi-cuadrado (χ^2), el índice de ajuste comparativo (CFI), y el error de aproximación de la media cuadrática (RMSEA). Generalmente se considera un buen ajuste con un CFI > .90 y RMSEA < .08, y excelente con CFI > .95 y RMSEA < .05 (Hu & Bentler, 1999). El CFI mide la mejora en el ajuste frente a un modelo nulo, mientras que el RMSEA es una medida de la cuantía del error. Se empleó el método de estimación de los mínimos cuadrados ponderados y varianza corregida (WLSMV), puesto que los ítems fueron medidos como variables categóricas (ordinales) y la literatura recomienda este método cuando las variables son de dicha naturaleza (Finney & DiStefano, 2006), además de no-normales.

Como medidas de fiabilidad de los factores se estimaron los valores tanto del alfa de Cronbach como del índice de confiabilidad compuesto

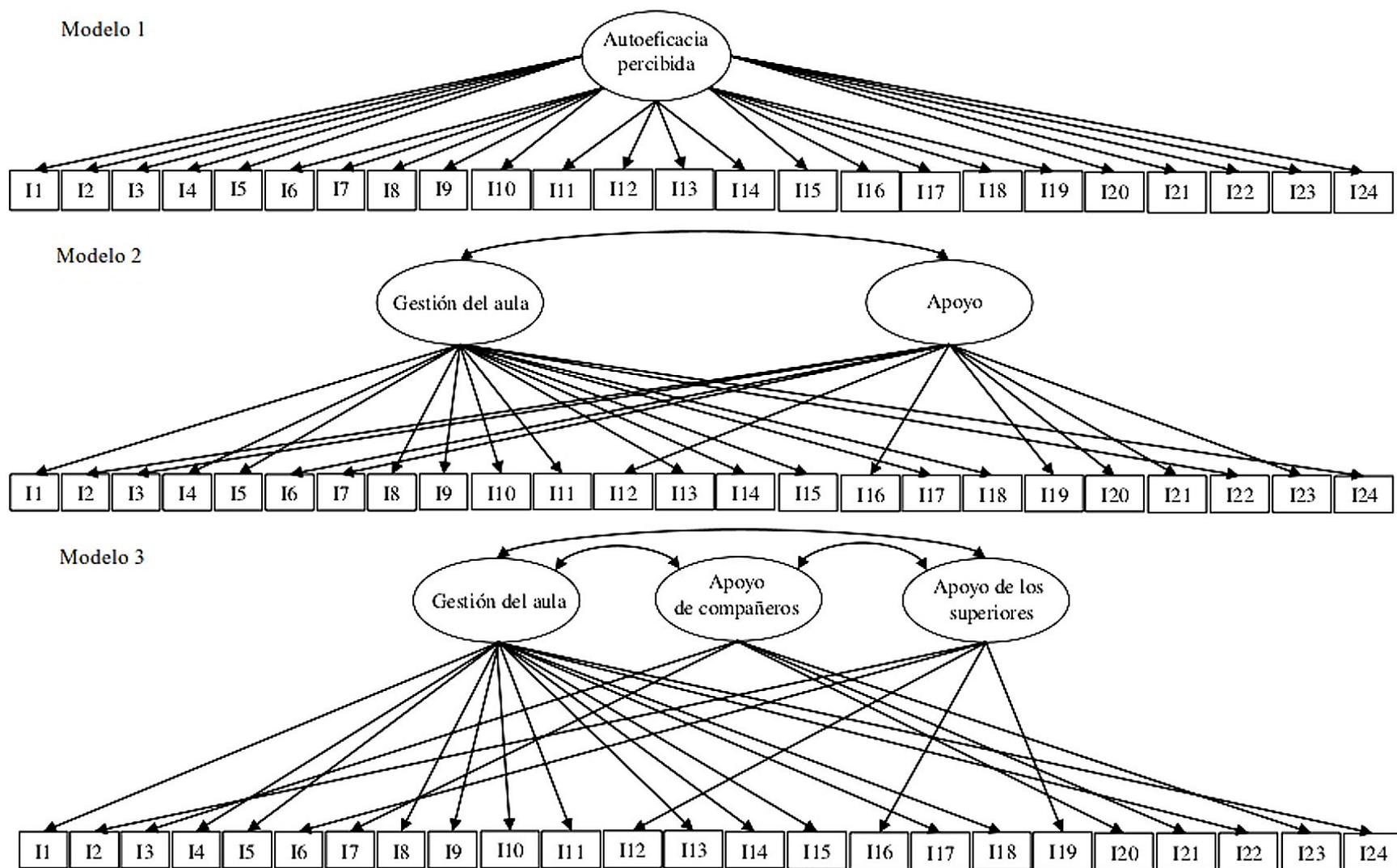


Figura 1

Modelos puestos a prueba en el estudio: a) un único factor de autoeficacia; b) dos factores: autoeficacia percibida en la gestión del aula y autoeficacia percibida en la obtención de apoyo; y c) tres factores: autoeficacia percibida en la gestión del aula, autoeficacia percibida en la obtención de apoyo de los compañeros, y autoeficacia percibida en la obtención de apoyo de los superiores.

Tabla 1

Índices de ajuste para los modelos testados.

Modelos	χ^2	<i>gl</i>	<i>p</i>	CFI	RMSEA	CI 90%
Modelo 1	1754.361	252	< .001	.943	.097	.093 - .102
Modelo 2	1505.824	251	< .001	.953	.089	.085 - .094
Modelo 3	1469.382	249	< .001	.954	.088	.084 - .093

Tabla 2

Correlaciones entre los factores de autoeficacia y los criterios de su red nomológica.

Autoeficacia	Expresión Generativa	Carencia Generativa	Motivación Intrínseca	Motivación Extrínseca	Motivación Intrínseca para Desarrollo Profesional
Autoeficacia General	.609**	.066	.585**	-.069	.239**
Gestión del Aula	.612**	-.083	.590**	-.109*	.241**
Apoyo Compañeros	.454**	-.098*	.437**	-.163**	.060
Apoyo Superiores	-.018	.297**	-.002	.150**	.115**

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

(CRI). Finalmente, se relacionaron los factores encontrados en la escala con una serie de criterios de su red nomológica, mediante coeficientes de correlación de Pearson. Los coeficientes alfa y las correlaciones de Pearson se calcularon en SPSS 21, mientras que los análisis factoriales confirmatorios y los CRI se calcularon en MPLUS 7 (Muthén & Muthén, 1998-2012).

Resultados

Estructura factorial

Los índices de ajuste de los tres modelos probados se muestran en la Tabla 1. Los tres modelos presentan un ajuste de bueno a excelente, pero el que mejores índices de ajuste presenta es el Modelo 3. El modelo de dos factores tiene un ajuste también excelente, sin embargo, la correlación entre los dos factores es muy alta, de .92 ($p <$

.01), lo que indica una baja validez discriminante entre los dos factores. Por su parte las correlaciones entre los factores del Modelo 3 son también muy elevadas. Así, se observa que *gestión del aula* se correlaciona un .90 ($p < .001$) con *obtención de apoyo de los compañeros*, un .89 ($p < .01$) con la *obtención de apoyo de superiores*, y *obtención de apoyo de los compañeros* y *obtención de apoyo de superiores* correlacionan .93 ($p < .001$).

Consistencia interna

Se calcularon el alfa de Cronbach y el CRI como estimadores de la consistencia interna de la escala. En el modelo de un solo factor, de autoeficacia docente general, el alfa fue de .91, mientras que en el modelo de tres factores los alfas fueron de .84 para *gestión del aula*, .786 para *obtención de apoyo de los compañeros* y .75 para *obtención*

de apoyo de los superiores. En cuanto al CRI, se obtuvo un valor de .95 para el modelo de un solo factor, valores de .91 para *gestión del aula*, .886 para *obtención de apoyo de los compañeros* y .86 para *obtención de apoyo de los superiores*.

Validez discriminante

Los resultados de relacionar los factores de la escala TISES con una serie de criterios de su red nomológica (*generatividad y motivación*) se presentan en la Tabla 2. Se ofrecen tanto las correlaciones de la solución de un factor como la de tres factores (los dos modelos mejores para representar los datos). Puede verse cómo la autoeficacia, como factor general, correlaciona alto y según lo esperado con *generatividad y motivación intrínseca*. Al dividir los ítems en las tres dimensiones del modelo de tres factores, puede apreciarse cómo estas no se comportan igual, más bien exhiben un comportamiento claramente idiosincrásico del factor *obtención de apoyo de los superiores*.

Discusión y conclusiones

Los resultados factoriales obtenidos no resultan tan claros como sería deseable. Por un lado, el modelo de tres factores propuesto teóricamente, y que encontró parcialmente apoyo en [García-Ros et al. \(2015\)](#), sí ajusta bien a los datos. Por otro lado, el modelo de un factor, mucho más parsimonioso, también presenta un buen ajuste. A favor del modelo de un factor se encuentra su parsimonia, pero sin embargo no es capaz de distinguir las obvias diferencias que se encuentran entre las dimensiones a priori en relación con otros constructos de su red nomológica. A favor del modelo de tres factores se encuentra que sus

dimensiones claramente presentan validez discriminante, que la capacidad de generalización se ve incrementada con respecto a la del modelo de un factor (es como se está empleando la escala en investigación), pero las correlaciones excesivamente altas entre estos factores hacen difícil negar la idea de que tienen un constructo general a la base, cuanto menos.

Por tanto, y en vista de los resultados obtenidos, hemos decidido tentativamente retener el modelo de tres factores como el que mejor representa los datos. Esta decisión se tomó teniendo en cuenta los índices de ajuste y la comparación de las correlaciones de los modelos de un factor y de tres factores con su red nomológica. A pesar de que los índices de ajuste para ambos modelos son buenos, en la tabla de correlaciones se puede ver como las correlaciones con la red nomológica del modelo de un factor son muy semejantes a las del primer factor del modelo de tres factores (*gestión del aula*). Esto indicaría que el segundo y el tercer factor del modelo de tres factores tienen una explicación idiosincrásica de la autoeficacia docente, y por tanto cabe tenerlos en cuenta. Esta es una forma de valorar la validez discriminante de la solución de tres factores con respecto a variables que puedan ser de interés para el investigador. Si los factores que se asumen con base en las respuestas de los profesores presentan correlaciones diferentes con otros constructos, es que pueden añadir algo específico a la explicación en el entorno educativo, lo que apoyaría la utilidad de la estructura más compleja. Estos resultados están en línea con los resultados hallados en el estudio original de la escala ([Brouwers & Tomic, 2001](#)), así como con los resultados obtenidos por [García-Ros et al. \(2015\)](#), quienes también defienden la estructura de tres factores. En el caso de la validación original, sin embargo, el ajuste no era plenamente satisfactorio de acuerdo a los criterios habituales. La validación en castellano de la esca-

la sí presentó un buen ajuste. Sin embargo, este buen ajuste solamente se produjo tras liberar cinco covarianzas de error, lo que no es deseable en el modelado estructural salvo que esté bien justificado teóricamente (Bollen, 1989) pues supone declarar que un residual (varianza no explicada) correlaciona con otro residual. Adicionalmente, el método de estimación empleado por ambos trabajos está superado por mejores estimaciones que corrijan la falta de normalidad y continuidad de las variables. Pese a ello, nuestros resultados se encuentran en la línea de los ya encontrados, aunque generan dudas sobre la validez discriminante de las dimensiones propuestas.

La existencia de medidas de autoeficacia específicas en la población docente es importante porque el sentimiento de autoeficacia de los docentes para motivar y promover el aprendizaje afecta al éxito académico que sus estudiantes son capaces de alcanzar (Bandura, 1993). Pudiendo medir los niveles de autoeficacia docente de una población específica se podrán desarrollar, en caso de que sea necesario, políticas e intervenciones encaminadas a fomentar el sentimiento de autoeficacia en los profesores, a fin de lograr un mayor bienestar para estos y mayores niveles de éxito académico en la población. Además del éxito académico, la autoeficacia docente se relaciona con una serie de variables en los estudiantes: motivación (Bergman et al., 1977), autoestima y actitudes prosociales (Borton, 1991; Cheung & Cheng, 1997). La autoeficacia docente se ha visto asociada también a una serie de variables organizacionales propias de los profesores: adopción de técnicas innovadoras (Fuchs, Fuchs, & Bishop, 1992), compromiso profesional (Coladarci, 1992), absentismo (Imants & Van Zoelen, 1995), estrés y *burnout* (Bliss & Finneran, 1991). Según Velázquez (2012), la autoeficacia también influye positivamente en la competencia, hecho que resulta relevante si tenemos en cuenta el papel que

juega la competencia en la labor de un maestro. Es por tanto importante tener a nuestra disposición herramientas adecuadas para medir la autoeficacia de los docentes, a fin de poder tener un control sobre la misma, pues esta influirá en una serie de factores cuyas consecuencias tienen un impacto directo sobre la calidad de la enseñanza.

En lo referente a potenciar el sentimiento de autoeficacia de los docentes, no deben olvidarse una serie de factores ajenos a ellos y que también influirán en el desarrollo de este sentimiento de autoeficacia: el nivel de rendimiento de los estudiantes, sus edades, el nivel de preparación del docente y el tamaño de la clase. Por tanto, una forma de potenciar la autoeficacia del docente será proporcionarle el medio y las herramientas necesarias. Si bien es cierto que variables tales como la edad de los estudiantes o el nivel del rendimiento de los mismos no pueden controlarse, otras como son el tamaño de la clase o el nivel de preparación del profesorado sí pueden controlarse. Por tanto, habrá que proporcionar al profesorado la preparación necesaria para su labor así como reducir, en la medida de lo posible, el tamaño de la clase donde el docente obra. En este aspecto, la posibilidad de tener una herramienta para medir el impacto que puedan tener estos cambios sin duda resulta de enorme utilidad, y de ahí deriva la necesidad así como la importancia de disponer de un instrumento que pueda medir la autoeficacia docente con certeza.

Con todo, resulta de vital importancia el estudio de la autoeficacia docente, así como su diferenciación en autoeficacia docente personal y general. La escala validada en la presente investigación contribuye al estudio de la autoeficacia docente personal. Las fortalezas de este estudio son: que constituye una contribución al cuerpo de conocimiento sobre el funcionamiento de la escala, que la muestra destaca por su naturaleza aleatoria y representativa, y que es la primera validación

de esta escala realizada en una muestra obtenida en la República Dominicana. Las limitaciones y, por tanto, las posibles mejoras serían: conseguir una muestra representativa, no sólo de docentes de distritos educativos de Santo Domingo, sino de la República Dominicana y, para valorar posibles invarianzas, sería conveniente recoger un mayor número de varones en la muestra.

Referencias

- Ashton, P. T., Olejnik, S., Crocker, L., & McAuliffe, M. (1982). Measurement problems in the study of teachers' sense of efficacy. *Annual Meeting of the American Educational Research Association*. New York.
- Ashton, P., Buhr, D., & Crocker, L. (1984). Teachers' sense of efficacy: A self- or norm-referenced construct. *Florida Journal of Educational Research*, 26(1), 29-41. Recuperado de <https://eric.ed.gov/?id=ED323221>
- Bandura, A. (1990). *Multidimensional scales of perceived academic efficacy*. Stanford, CA: Stanford University.
- Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist*, 28(2), 117-148. doi: 10.1207/s15326985ep2802_3
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.
- Bandura, A. (2000). Cultivate self-efficacy for personal and organizational effectiveness. En E. A. Locke (Ed.), *Handbook of Principles of Organization Behavior* (pp. 120-136). Oxford: Blackwell.
- Bergman, P., McLaughlin, M., Bass, M., Pauly, E., & Zellman, G. (1977). *Federal programs supporting educational change: Vol. VII. Factors affecting implementation and continuation*. Santa Monica, CA: RAND.
- Bliss, J., & Finneran, R. (1991). Effects of school climate and teacher efficacy on teacher stress. *Annual Meeting of the American Educational Research Association*. Chicago.
- Bollen, K. A. (1989). *Structural Equation Modeling with Latent Variables*. Mahwah, NJ: LEA.
- Borton, W. M. (1991). Empowering teachers and students in a restructuring school: A teacher efficacy interaction model and the effect on reading outcomes. *Annual Meeting of the American Educational Research Association*. Chicago.
- Brouwers, A. (2000). *Teacher burnout and self-efficacy: An interpersonal approach*. Heerlen, Netherlands: Open University.
- Brouwers, A., & Tomic, W. (2001). The factorial validity of scores on the teacher interpersonal self-efficacy scale. *Educational and Psychological Measurement*, 61(3), 433-445. doi: 10.1177/00131640121971301
- Brouwers, A., Tomic, W., & Stijnen, S. (2002). A confirmatory factor analysis of scores on the teacher efficacy scale. *Swiss Journal of Psychology*, 61(4), 211-219. doi: 10.1024/1421-0185.61.4.211
- Chacón, C. (2006). Las creencias de autoeficacia: Un aporte para la formación del docente de inglés. *Acción Pedagógica*, 15(1), 44-54. Recuperado de <http://www.saber.ula.ve/handle/123456789/17262>
- Cheung, W. M., & Cheng, Y. C. (1997). A multi-level analysis of teachers' self-belief and behavior, and students' educational outcomes. *Annual Meeting of the American Educational Research Association*. Chicago.
- Coladarci, T. (1992). Teachers' sense of efficacy and commitment to teaching. *Journal of Experimental Education*, 60(4), 323-337. doi: 10.1080/00220973.1992.9943869
- Emmer, E. T., & Hickman, J. (1991). Teacher efficacy in classroom management and discipline. *Educational and Psychological Measurement*, 51(3), 755-765. doi: 10.1177/0013164491513027
- Fernández, B. (2012). *Un análisis multidimensional del síndrome de burnout en profesorado de conservatorios y enseñanza secundaria*. (Tesis de Doctorado). Universidad de Valencia, España.
- Fernández, J. (1987). *Escala de Motivaciones Psicosociales*.

Madrid: Tea.

- Finney, S. J., & DiStefano, C. (2006). Non-normal and categorical data in structural equation modeling. En G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 269-314). Greenwich, CT: Information Age.
- Fuchs, L. S., Fuchs, D., & Bishop, N. (1992). Instructional adaptation for students at risk. *Journal of Educational Research*, 86(2), 70-84. doi: 10.1080/00220671.1992.9941143
- García-Ros, R., Fuentes, M. C., & Fernández, B. (2015). Teachers' interpersonal self-efficacy: Evaluation and predictive capacity of teacher burnout. *Electronic Journal of Research in Educational Psychology*, 13(3), 483-502. doi: 10.14204/ejrep.37.14105
- Gibson, S., & Dembo, M. H. (1984). Teacher efficacy: A construct validation. *Journal of Educational Psychology*, 76(4), 569-582. doi: 10.1037/0022-0663.76.4.569
- Gonzalo-Delgado, M., & León del Barco, B. (1999). La promoción de la autoeficacia en el docente universitario. *Revista Electrónica Interuniversitaria de Formación del Profesorado*, 2(1), 79-90. Recuperado de <https://dialnet.unirioja.es/servlet/articulo?codigo=2794996>
- Guskey, T. R. (1981). Measurement of responsibility teachers assume for academic successes and failures in the classroom. *Journal of Teacher Education*, 32(3), 44-51. doi: 10.1177/002248718103200310
- Henson, R. K., Kogan, L. R., & Vacha-Haase, T. (2001). A reliability generalization study of the Teacher Efficacy Scale and related instruments. *Educational and Psychological Measurement*, 61(3), 404-420. doi: 10.1177/00131640121971284
- Hu, L., & Bentler, P. M. (1999). Cut-off criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1-55. doi: 10.1080/10705519909540118
- Imants, J., & Van Zoelen, A. (1995). Teachers' sickness absence in primary schools, school climate and teachers' sense of efficacy. *School Organization*, 15(1), 77-86. doi: 10.1080/0260136950150109
- McAdams, D. P., & St. Aubin, E. (1992). A theory of generativity and its assessment through self-report, behavioral acts, and narrative themes in autobiography. *Journal of Personality and Social Psychology*, 62(6), 1003-1015. doi: 10.1037/0022-3514.62.6.1003
- Muthén, L. K., & Muthén, B. O. (1998-2012). *Mplus User's Guide* (7^a ed.). Los Angeles, CA: Muthén & Muthén.
- Prieto, L. (2007). *Autoeficacia del profesor universitario. Eficacia percibida y práctica docente*. Madrid: Narcea.
- Riggs, I., & Enochs, L. (1990). Toward the development of an elementary teacher's science teaching efficacy belief instrument. *Science Education*, 74(6), 625-638. doi: 10.1002/sce.3730740605
- Rose, J. S., & Medway, F. J. (1981). Measurement of teachers' beliefs in their control over student outcome. *Journal of Educational Research*, 74(3), 185-190. doi: 10.1080/00220671.1981.10885308
- Tanaka, J. S. (1993). Multifaceted conceptions of fit in structural equation models. En K. A. Bollen & J. S. Long (Eds.), *Testing Structural Equation Models* (pp. 10-39). Newbury Park, CA: Sage.
- Velásquez-Fernández, A. (2012). Revisión histórico-conceptual del concepto de autoeficacia. *Revista Pequeño*, 2(1), 148-160. Recuperado de <http://revistas.ubiobio.cl/index.php/RP/article/view/1838>

ANEXO I
Ítems de la escala Teacher Interpersonal Self-Efficacy Scale.

1	Si un alumno interrumpe la clase, soy capaz de reconducirlo rápidamente.
2	Soy capaz de acercarme a los directivos si quiero hablar de problemas con el trabajo.
3	Confío en que, si es necesario, puedo pedir consejo a mis compañeros.
4	Son muy pocos los alumnos que no puedo soportar.
5	Puedo sobrellevar los estudiantes más complicados.
6	Cuando es necesario, soy capaz de sacar a colación problemas con los directivos.
7	Siempre puedo encontrar compañeros con los que puedo hablar de problemas en el trabajo.
8	Puedo tomar medidas adecuadas para mantener actividades funcionando de manera adecuada.
9	Puedo decirles a los alumnos que me tomo muy en serio conseguir un trato apropiado.
10	No siempre soy capaz de ejecutar varias actividades a la vez.
11	Puedo manejar mi clase bastante bien.
12	Confío en que, si es necesario, puedo conseguir que los directivos me ayuden.
13	Puedo mantener a los alumnos desafiantes implicados en la clase.
14	Soy siempre capaz de expresar de forma clara mis expectativas a los alumnos.
15	Soy capaz de responder de manera adecuada a estudiantes desafiantes.
16	Cuando es necesario soy capaz de conseguir que los directivos me apoyen.
17	Puedo evitar que unos pocos alumnos desafiantes me arruinen una clase entera.
18	Si los alumnos paran de trabajar puedo devolverles a la tarea.
19	Confío en que si es necesario puedo pedir consejo a los directivos.
20	Si me enfrento a un problema en el que mis compañeros me pueden ayudar, soy capaz de acercarme a ellos para ello.
21	Cuando es necesario soy capaz de pedir ayuda a un compañero.
22	Sé qué reglas son las apropiadas para mis alumnos.
23	Soy capaz de acercarme a mis compañeros si quiero comentar problemas sobre el trabajo.
24	Soy capaz de empezar el curso de manera que los alumnos aprenderán a comportarse bien.

Evaluación de un Instrumento de Apoyo Institucional en Casos de Maltrato Infantil

Analysis of an Instrument for Assessing Institutional Support in Cases of Child Maltreatment

Valeria Arredondo *¹, Carolina Saavedra¹, Cristóbal Guerra¹

1 - Centro de Estudios en Infancia, Adolescencia y Familia, ONG Paicabi, Chile.

Introducción
Método
Resultados
Discusión
Referencias

Recibido: 22/08/2017 **Revisado:** 27/09/2017 **Aceptado:** 11/10/2017

Resumen

En Chile no existen suficientes instrumentos que evalúen de forma confiable y válida los factores involucrados en el maltrato infantil. En particular hay carencia de instrumentos que se centren en el rol que juegan las instituciones gubernamentales y no gubernamentales en la protección y apoyo a niños maltratados. Por esta razón el objetivo del estudio fue evaluar las propiedades psicométricas de un instrumento para evaluar el apoyo social institucional en casos de maltrato infantil grave (I-APSI). Se consideraron los datos de 544 niños, niñas y adolescentes atendidos en centros especializados en Chile. Se aplicaron el I-APSI y dos instrumentos que evalúan constructos relacionados. Los resultados apoyan la validez factorial del instrumento, su validez convergente y su confiabilidad. Pese a lo alentador de estos resultados, se debe considerar que el estudio es preliminar y se debe seguir profundizando en las propiedades psicométricas del instrumento.

Palabras clave: *maltrato infantil, apoyo social, protección, confiabilidad, validez*

Abstract

There are not enough reliable and valid instruments to evaluate the factors involved in child maltreatment in Chile. In particular, there is a lack of instruments that focus on the role of governmental and non-governmental institutions in protecting and supporting abused children. For this reason, the objective of the study was to evaluate the psychometric properties of an instrument that assesses institutional social support in cases of severe child maltreatment (I-APSI). Data from 544 children and adolescents that were receiving treatment in specialized centers in Chile was contemplated. I-APSI and two instruments assessing related constructs were applied. The results support the factorial validity of the instrument, its convergent validity and its reliability. These results should be considered as preliminary, and the psychometric properties of the instrument should be further investigated in the future.

Key words: *children maltreatment, social support, protection, reliability, validity*

*Correspondencia a: valeria.arredondo@paicabi.cl

Introducción

El *maltrato infantil*, dentro del cual se incluyen la negligencia, el maltrato psicológico, el maltrato físico y el abuso sexual, es una de las formas más graves de vulneración de derechos contra menores de 18 años (UNICEF, 1989). Se considera que sus consecuencias son de alto riesgo para el desarrollo y bienestar de los niños y niñas, dado que quienes lo experimentan evidencian dificultades en casi todas las dimensiones evolutivas (Morelato, 2011), entre las que se encuentran la presencia de depresión, baja autoestima, agresividad, desesperanza, escasas habilidades de afrontamiento, dificultades en el control de impulsos y regulación emocional y peor desempeño académico en comparación con los niños que no han sido víctimas de malos tratos en su vida. Debido justamente a las serias consecuencias emocionales, sociales, cognitivas y comportamentales que genera en sus víctimas, el maltrato infantil es un hecho rechazado socialmente en la mayoría de los países en forma cada vez más clara y sostenida (Barudy, 1998), aspecto que se instala en la agenda internacional a partir de la Convención Internacional de los Derechos del Niño (UNICEF, 1989), que en su articulado determina el rechazo al trato duro o violento hacia los niños, comprometiendo a los estados a generar las políticas públicas necesarias para su implementación. En coherencia con esta visibilidad social del maltrato infantil, en 1990 el Estado de Chile ratifica dicha Convención y, mediante este acto, se compromete a velar por el respeto a los derechos de los menores de 18 años en el país. Este compromiso abarca a todas las instituciones del Estado, quienes se ven obligadas a considerar dentro de sus lineamientos una posición coherente con la promoción de derechos de la infancia, la protección de sus intereses y necesidades, y la reparación integral tras la vulneración de sus derechos fundamentales.

En este escenario, si bien es posible considerar que el Estado de Chile ha avanzado en este cambio de paradigma hacia la protección especializada de la infancia tras la Convención Internacional de los Derechos del Niño, a través de la generación de políticas públicas focalizadas, la creación de centros de atención especializados para cubrir los procesos reparatorios tras la identificación de la vulneración de derechos y avances en la legislación en infancia, adolescencia y familia (Guerra & Arredondo, 2017), también es posible indicar que las tasas de maltrato infantil en Chile siguen siendo elevadas pese a estos esfuerzos. En este sentido, estudios chilenos estiman que el 75.1% de los niños y niñas del país habrían sido maltratados al menos alguna vez en su vida (Ministerio del Interior, 2013). Por su parte, el Cuarto Estudio de Maltrato Infantil, desarrollado por UNICEF en Chile en el año 2012, basado en la aplicación de cuestionarios autoadministrados a niños y niñas en establecimientos educacionales, al comparar los resultados de este estudio con los obtenidos en 1994, 2000 y 2006; señala que el 71% de los niños y niñas recibe algún tipo de violencia de parte de su madre y/o padre, un 51.5% sufre algún tipo de violencia física y el 25.9% de los niños y niñas sufre violencia física grave. Cabe destacar que, si bien los datos ilustran una disminución de la violencia general hacia los niños y niñas entre los años 1994 y 2012, la violencia física grave se ha mantenido sin modificaciones (UNICEF, 2012). Por su parte, un reciente estudio realizado con adolescentes en Chile indica que el 89% de un total de 706 participantes informa que ha sufrido algún tipo de victimización o maltrato a lo largo de su vida (Pinto-Cortez & Venegas-Sanhueza, 2015). En este estudio se destaca la experiencia de hasta seis acontecimientos distintos de maltrato en el 68.1% de los casos, en tanto que el 30.3% declara haber experimentado siete o más eventos de maltrato o violencia. Estos

datos indican que el maltrato ocurre más en episodios cotidianos repetidos que en eventos aislados, aspecto que resulta fundamental de considerar a la hora de generar un marco comprensivo del fenómeno para su enfrentamiento e intervención.

Ahora bien, Cichetti, Rogosch, Lynch y Holt (1993) sostienen que no todos los niños y niñas son afectados del mismo modo por estas experiencias de maltrato infantil, pues estas dependen de su edad y período de desarrollo, así como de los contextos y formas de maltrato, es decir que cada experiencia está configurada en un entramado de condiciones que se asocian a su impacto específico. En esta línea, si consideramos el maltrato infantil como una experiencia adversa que implica un riesgo de sufrir distintos tipos de consecuencias, Luthar, Cicchetti y Becker, (2000) plantean que pueden existir variaciones en el funcionamiento y desarrollo de los individuos expuestos a este tipo de experiencias, enfatizando que podrían manifestar competencia en algunas áreas y exhibir algunas dificultades en otras en un patrón más bien variable. Lo anteriormente expuesto es un elemento crucial que debe ser considerado por los distintos servicios que brindan atención y asistencia a niños, niñas y adolescentes que han experimentado maltrato, dado que a partir de este tipo de consideraciones se puede diseñar una ayuda atinente a las necesidades y posibilidades de cada niño o niña víctima, considerando sus particularidades. En la temática del maltrato infantil, uno de los aspectos que es importante integrar es el *apoyo social*, dado que constituye una variable destacada en la protección contra el maltrato infantil y también está asociado a su impacto, en la medida que tendría una función mediadora o moderadora entre la experiencia de maltrato y el desarrollo de sintomatología psicológica (Evans, Steel, & DiLillo, 2013; Guerra, Ocaranza, & Weinberger, 2016; Guerra, Pereda, Guilera, & Abad, 2016; Pereda-Beltran, 2009;

Turner, Shattuck, Finkelhor, & Hamby, 2015), debido a lo cual su consideración dentro de los procesos de intervención resulta relevante.

El apoyo social puede ser entendido como el proceso interaccional por el cual los recursos en la estructura social (como la comunidad y redes sociales) hacen posible la satisfacción de distintos tipos de necesidades en situaciones cotidianas y de crisis (Gracia & Musitu, 1993). Dentro de las posibilidades de análisis del apoyo social, se encuentra la consideración de distintas fuentes de origen y mantención, tales como la red de relaciones sociales donde se encuentran las relaciones íntimas, con vecinos, amigos, compañeros de trabajo, entre otros, y las organizaciones y servicios formales e informales, que comprenden las asociaciones cívicas, servicios sociales y de apoyo profesional, entre otras (Gracia & Musitu, 1993). Otro enfoque de análisis de esta variable es el de la perspectiva relacional vinculada con el niño o niña. Así, Luthar y Goldstein (2004) señalan que los padres de los niños y niñas, así como otros adultos presentes en su contexto de vida, pueden proveer y proveer un valioso soporte frente a las situaciones de adversidad. Por ello, se debe alentar la relación con adultos de confianza, lo cual contribuiría a disminuir las consecuencias de los malos tratos en los niños y adolescentes; este aspecto debe ser considerado dentro de la intervención. Otros han considerado lo opuesto, es decir que más que disminuir las consecuencias de los malos tratos, la relación con adultos de confianza es un factor que contribuye al bienestar por sí solo, lo que le otorga valor en el desarrollo de las necesidades de afiliación, pertenencia, respeto, y reconocimiento social (Gracia & Musitu, 1993). En cualquier caso, su consideración como variable relevante dentro de la evaluación e intervención en maltrato infantil queda establecida.

No obstante, una de las dificultades para los centros de atención especializada en atención a

víctimas de maltrato infantil grave es que no se cuenta con instrumentos confiables y válidos para evaluar los avances de las intervenciones y las condiciones de esta variable necesarias para abordar la problemática en términos reparatorios (Arredondo, Saavedra, & Guerra, 2017). Desde este punto de vista, se considera pertinente aumentar los esfuerzos por diseñar o adaptar instrumentos que puedan ser utilizados en Chile, de manera fiable y que resulten ser una guía para el diseño de la intervención.

Debido a lo anterior, la ONG Paicabi (2012) diseñó un instrumento para evaluar una dimensión específica del apoyo social institucional, la visibilización del niño y de su familia. Este apoyo es otorgado por organismos sociales ante situaciones de maltrato infantil grave (I-APSI). La relevancia de evaluar el apoyo social institucional está situada en dos aspectos que convergen. El primero, un nivel netamente ético-político vinculado con el seguimiento de los postulados de la Convención de los Derechos del Niño (UNICEF, 1989) que, como ya se ha dicho, obliga a todas las instituciones de los estados adscritos a velar por la protección de los niños y niñas, y ejercer un rol fiscalizador y reparador ante una situación de vulneración de derechos como lo es el maltrato infantil. En este sentido, la red institucional formal comprende una red de recursos para la superación del maltrato como problema social. El segundo, un nivel pragmático asociado con la evidencia empírica que pone en relieve la importancia del apoyo social. En este sentido, se decidió generar un instrumento que evalúe el apoyo de las redes sociales institucionales debido a la consideración de que el apoyo social es una de las variables que más frecuentemente se destacan en la literatura como factor protector ante situaciones y consecuencias del maltrato (Evans et al., 2013; Guerra, Ocaranza, et al., 2016; Guerra, Pereda, et al., 2016; Turner et al., 2015). De hecho, se plan-

tea que los soportes extrafamiliares constituyen un factor asociado con la resiliencia en niños y niñas que han sufrido violencia, por lo cual su consideración en el análisis de las victimizaciones infantiles tiene alcances evaluativos e interventivos (Morelato, 2011). Este factor comprende los entornos que brinden sostén desde el contexto comunitario del niño o niña (escolar, familiar, religioso, etc.): actividades extracurriculares, experiencias positivas en la escuela y la posibilidad de tener un entorno estable y/o una comunidad religiosa de referencia. Estos entornos promueven el desarrollo de factores de resiliencia en niños y adolescentes. Es importante tener en cuenta que este factor se combina e interactúa con ciertos aspectos individuales tales como adecuadas habilidades cognitivas y autoestima. En este sentido, es importante señalar que los aspectos individuales asociados con la resiliencia pueden haberse desarrollado justamente gracias a la presencia de alguna forma de protección en los contextos de cuidado (familias sustitutas o un buen ambiente escolar, por ejemplo), especialmente durante la infancia temprana, aspecto que ha sido relevado por distintos autores (Egeland, Carlson, & Sroufe, 1993; Valentine & Feinahuer, 1993).

Considerando todo lo anterior, el I-APSI fue elaborado por la ONG Paicabi a través de la colaboración de distintos profesionales expertos (trabajadores sociales y psicólogos) que atienden a niños, niñas y adolescentes que han sufrido maltrato infantil grave. Su finalidad es constituir una herramienta útil para la valoración de los factores de apoyo social institucional vinculados con la visibilización del niño, niña o adolescente y así contribuir a determinar los objetivos de la intervención. La versión inicial fue creada en el año 2003 por un comité de expertos de la misma institución y contó con un total de 12 ítems. Desde ese año se han realizado 3 pilotajes que culminaron en el año 2012 con la versión final, que cuenta

con un total de 10 ítems. El instrumento es heteroaplicado por los profesionales a cargo de la atención directa de los niños y niñas participantes de los centros especializados. Para poder aplicarlo de forma adecuada es necesario que los profesionales se capaciten en ello basándose en el manual creado para tal efecto (Paicabi, 2012).

Si bien desde el año 2012 a la fecha el I-APSI ha sido aplicado por los profesionales responsables del proceso de atención especializada al interior de la ONG, aún no se han estudiado las propiedades psicométricas de este instrumento. Por esta razón, el objetivo del estudio es evaluar las propiedades psicométricas del I-APSI, considerando los datos tomados en función de niños, niñas y adolescentes chilenos. En particular se evalúa la estructura factorial del instrumento, confiabilidad y validez convergente mediante el análisis de la relación con otras dos medidas que evalúan algunas variaciones del apoyo del entorno social (específicamente el reconocimiento y protección de la red institucional ante situaciones de maltrato). Se espera que el instrumento demuestre una adecuada consistencia interna y una estructura factorial compatible con la teoría y que sus puntuaciones se asocien directamente con las otras medidas de apoyo social.

Método

Participantes

En este estudio se evaluó el apoyo social institucional recibido por niños, niñas y adolescentes que se encontraban en intervención por maltrato infantil grave en centros especializados. Para ello, se consideraron los datos de las fichas clínicas de 544 niños, niñas y adolescentes atendidos en once centros de intervención especializada en maltrato infantil grave de la ONG Paicabi, de la V° Región en Chile. Únicamente se consideraron los

casos de aquellos niños, niñas y adolescentes de los cuales se tenía un completo diagnóstico de los factores involucrados en el presente estudio.

Del total de casos analizados, 339 (62.3%) eran de género femenino y 205 (37.7%) eran de género masculino. Las edades de los participantes fluctuaron entre 1 y 17 años ($M = 10.13$; $DE = 3.99$). Del total de casos, 288 (52.9%) habían sido víctimas de delitos sexuales, 109 (20%) habían sido testigos de violencia intrafamiliar, 78 (14.3%) habían sufrido maltrato físico y 67 (12.3%) maltrato psicológico.

Instrumentos

En este estudio se utilizaron 3 instrumentos. El primero de ellos es el instrumento para evaluar el apoyo social institucional en casos de maltrato infantil grave, cuyo análisis psicométrico es el objetivo del estudio. Los otros dos instrumentos han sido incluidos como medidas de constructos asociados, relevantes para el proceso de validez convergente del I-APSI.

Instrumento para evaluar el apoyo social institucional en casos de maltrato infantil grave (I-APSI). El instrumento debe ser respondido por el equipo profesional tras la evaluación diagnóstica de cada niño, niña o adolescente en base a su protocolo de aplicación (Paicabi, 2012). La versión aplicada en este estudio posee 10 ítems. El formato de respuesta de cada ítem ofrece 3 alternativas, donde: 1 = *Baja presencia del indicador*, 2 = *Presencia moderada del indicador* y 3 = *Alta presencia del indicador*. Las propiedades psicométricas del instrumento se describen en los siguientes apartados.

Factor contextual del instrumento para evaluar el reconocimiento de situaciones de maltrato infantil I- REC (Paicabi, 2012). Se trata de

una prueba heteroaplicada de 7 ítems que evalúa en qué medida la red de apoyo institucional que rodea al niño es capaz de reconocer el carácter abusivo de situaciones de maltrato infantil grave (ejemplo de indicador: la red focal identifica los derechos vulnerados del niño, niña o adolescente). El formato de respuesta de cada ítem ofrece 3 alternativas, donde: 1 = *Baja presencia del indicador*, 2 = *Presencia moderada del indicador* y 3 = *Alta presencia del indicador*. Los puntajes posibles varían entre los 7 y 21 puntos (a mayor puntuación, mayor reconocimiento de situaciones de maltrato infantil). El instrumento posee adecuadas propiedades psicométricas en población chilena (Arredondo, Saavedra, & Guerra, 2017). En este estudio el instrumento obtuvo una buena confiabilidad ($\alpha = .88$).

Factor contextual del instrumento para evaluar indicadores de protección en el contexto de maltrato infantil grave I-PROT (Paicabi, 2012; Arredondo-Ossandon, Saavedra-Inostroza, & Guerra-Vio, 2016). Se trata de un instrumento heteroaplicado de 5 ítems que evalúa los recursos protectores del contexto social institucional (p. ej., presencia de redes focales que visibilicen el maltrato). El formato de respuesta de cada ítem ofrece 3 alternativas, donde: 1 = *Baja presencia del indicador*, 2 = *Presencia moderada del indicador* y 3 = *Alta presencia del indicador*. Los puntajes posibles varían entre los 5 y 15 puntos (a mayor puntuación mayor presencia de indicadores protectores). En este estudio el instrumento obtuvo una buena confiabilidad ($\alpha = .86$).

Procedimiento

Previo a su implementación, el proyecto fue aprobado por un comité de ética institucional compuesto por profesionales especializados en el área y externos al equipo de investigación.

Una vez seleccionados los niños, niñas y adolescentes participantes del estudio, los profesionales respondieron los instrumentos ya especificados. Es necesario señalar que todos los profesionales fueron capacitados sobre cómo responder los instrumentos de este estudio. En todo momento se guardó reserva de los datos de identificación de cada participante.

Análisis de datos

En primer lugar, los 10 ítems iniciales del I-APSI fueron sometidos a un análisis factorial exploratorio de ejes principales con rotación oblimin. Luego, se evaluó la confiabilidad de los factores resultantes mediante el cálculo del coeficiente alfa de Cronbach y alfa ordinal (Domínguez-Lara, en prensa). En tercer lugar, se evaluó la validez convergente del instrumento mediante el análisis de la correlación entre sus puntuaciones y las de las dimensiones contextuales del I-REC y del I-PROT. Este análisis fue realizado con el coeficiente *rho* de Spearman. Finalmente, se calcularon puntajes cuartiles para interpretar provisionalmente los resultados de futuras aplicaciones.

Resultados

Estructura factorial y confiabilidad

Previo al análisis factorial se analizó el coeficiente de significación de Kaiser Mayer (KMO = .80) y el índice del test de esfericidad de Barlett ($\chi^2 = 3811,329$; $p < .01$). Ambos análisis sugieren que los ítems están relacionados entre sí, lo que permite su factorización (Vivanco, 1999). El resultado del análisis factorial arrojó 2 factores con autovalores mayores a 1. No obstante, se descartó esta solución ya que el factor 1 estaría conforma-

Tabla 1

Cargas factoriales en la solución de dos factores y cargas factoriales, relaciones ítem-test y alfa si se elimina el ítem en la solución de un factor.

	Solución de dos factores		Solución de un factor		
	Cargas en Factor 1	Cargas en Factor 2	Cargas en Factor único	Relación ítems test	α si se elimina el ítem
Visibilidad del niño(a) o joven en la Red Educativa	.55		.74	.70	.90
Visibilidad del niño(a) o joven en la Red de Salud	.46	.39	.75	.71	.90
Visibilidad del niño(a) o joven en la Red Institucional	.74		.79	.75	.90
Visibilidad del niño(a) o joven en la Red Judicial	.84		.63	.59	.91
Visibilidad del niño(a) o joven en la Red Comunitaria		.83	.61	.60	.91
Visibilidad de la familia en la Red Educativa	.54	.31	.77	.73	.90
Visibilidad de la familia en la Red de Salud	.53	.38	.81	.77	.90
Visibilidad de la familia en la Red Institucional	.70		.77	.72	.90
Visibilidad de la familia en la Red Judicial	.86		.71	.67	.91
Visibilidad de la familia en la Red Comunitaria		.77	.64	.59	.91

Nota. Sólo se incluyeron cargas factoriales $\geq .3$.

do por 8 ítems y el factor 2 sólo estaría compuesto de 2 ítems, lo que no es recomendable (Kline, 2005). Además, tres de los ítems presentaban cargas factoriales mayores a .3 en más de un factor, lo que sugiere que podría obtenerse una solución de mayor parsimonia (ver Tabla 1).

Efectivamente, al repetir el análisis prefijando la obtención de un factor único, todos los ítems presentan cargas factoriales altas en el factor (sobre .6). El factor único permite explicar el 57.1% de la variabilidad de los ítems. Además se aprecia que el factor único presenta adecuada confiabilidad (α de Cronbach = .91; α ordinal = .92). Todos los ítems se encuentran asociados con la escala total y la eliminación de ninguno de ellos aumentaría la consistencia interna de la escala.

Validez convergente y estadísticos descriptivos del I-APSI

Como se puede ver en la Tabla 2, la puntuación del I-APSI está directamente relacionada con las puntuaciones de la dimensión contextual del I-REC ($M = 16.48$; $DE = 3.50$). Como ya se ha señalado, el I-REC evalúa en qué medida la red de apoyo institucional que rodea al niño es capaz de reconocer el carácter abusivo de situaciones de maltrato infantil grave. De esta manera, y de acuerdo con lo esperado, cuanto más visibiliza la red institucional al niño o a su familia mayor tiende a ser el reconocimiento por parte de esta red de las situaciones de maltrato infantil grave.

En la misma línea, la puntuación del I-APSI también se relaciona directamente con las pun-

tuaciones de la dimensión contextual del I-PROT ($M = 11.92$; $DE = 2.62$). Esto da cuenta de que a mayor apoyo de las redes sociales institucionales que rodean al niño, niña o adolescente hay mayor disponibilidad de recursos proteccionales en ese contexto, lo que es compatible con las hipótesis planteadas.

Tabla 2

Relación de las puntuaciones del I-APSI con los constructos asociados y estadísticos descriptivos.

	I-APSI
I-PROT Contextual	$\rho .55^{**}$
I-REC Contextual	$\rho .51^{**}$
Media (DE)	24.06 (4.75)
Asimetría	-.517
Curtosis	-.072
Valor mínimo	10
Cuartil 1	20
Cuartil 2	24
Cuartil 3	28
Valor Máximo	30

Nota. Correlaciones rho de Spearman: $**p < .01$.

La Tabla 2 también muestra datos descriptivos de las puntuaciones del instrumento. Específicamente, la media, desviación típica, valores mínimos, cuartil 1, cuartil 2, cuartil 3, valor máximo, asimetría y curtosis. Estos valores pueden ser usados como referencias para la interpretación de los resultados del instrumento en futuras aplicaciones. A la hora de interpretar las puntuaciones de esta muestra es importante considerar que los valores posibles oscilan entre los 10 y 30 puntos.

Discusión

El objetivo de este estudio fue evaluar las propiedades psicométricas del I-APSI que, si bien

está siendo utilizado en Chile por la ONG Paicabi desde el año 2012, no contaba con respaldo empírico de su confiabilidad ni de su validez. De este modo, la relevancia de la investigación radica en la evidencia otorgada respecto de las garantías psicométricas que posee el instrumento, dado que con él se toman importantes decisiones respecto a las intervenciones de niños, niñas y adolescentes cuyos derechos han sido gravemente vulnerados. Asimismo, la plataforma de intervención en maltrato infantil a nivel nacional carece de este tipo de instrumentos a pesar de que el maltrato infantil es una dimensión de trabajo interventivo definida desde la política pública en infancia presente en todo Chile por normativa técnica nacional (SENAME, 2011), por lo cual se supone que el instrumento resultará útil en este contexto.

En particular, el I-APSI evalúa un constructo escasamente medido por otros instrumentos que evalúan el apoyo social. La mayoría de los instrumentos se centran en la percepción subjetiva de apoyo social del niño, niña o adolescente, sin considerar una medida externa del apoyo institucional (Terol-Cantero et al., 2004). Los instrumentos que teóricamente presentan mayor convergencia con el I-APSI son el I-REC y el I-PROT en sus dimensiones contextuales (Paicabi, 2012) que, aunque evalúan otras dimensiones más específicas del apoyo institucional (reconocimiento de situaciones abusivas y prácticas protectoras respectivamente), no incorporan reactivos referidos a un elemento más elemental como es la visibilización del niño y de su familia.

Se considera que esta omisión en la mayoría de los instrumentos de evaluación del apoyo social genera un importante vacío, especialmente en el contexto del maltrato infantil grave, en el que las condiciones proteccionales de la vida del niño o niña pasan muchas veces por la acción institucional coherente y ajustada a sus necesidades, aspecto fundamental en la intervención especiali-

zada (Barudy, 1998, 2001). Respecto a este punto, las normativas técnicas a nivel nacional para los programas de atención especializada en maltrato infantil desde el Estado de Chile incorporan esta dimensión tanto en la evaluación como en la acción interventiva (SENAME, 2011); sin embargo, no proponen instrumentos específicos para abordar esta dimensión, lo que crea el vacío que la presente investigación pretendió abordar.

Desde la perspectiva ecológica, se puede entender que en la mantención de las prácticas abusivas en contra de niños, niñas o adolescentes operan una serie de factores individuales, familiares y contextuales para los que las instituciones sociales, especialmente las que provienen del estado, juegan un rol importante (Belsky, 1980; Bronfenbrenner, 1987). De hecho, la Convención de los Derechos del Niño (UNICEF, 1989) apunta precisamente a eso, a que las instituciones del Estado se comprometan por velar por los derechos de los menores de edad, de manera que se disminuya el maltrato en todos sus niveles. Ahora bien, entendiendo que en el abordaje de la problemática del maltrato infantil la primera acción por realizar es tomar una medida protectora que ponga al niño fuera de peligro, estas acciones no son suficientes para la superación de las consecuencias del maltrato. Se ha observado que a medida que transcurre el tiempo suelen manifestarse consecuencias del maltrato a pesar de que los niños ya no están en peligro (Cicchetti & Rogosch, 1997; Flores, Cicchetti, & Rogosch, 2005), probablemente porque un factor de protección no es siempre un factor de resiliencia (Cyrułnik, 2003). Por ello, se torna necesario evaluar y consecuentemente intervenir en el apoyo social institucional promoviendo el fortalecimiento de las redes sociales (Morelato, 2011).

Desde este punto de vista, para la ONG Paicabi cobra especial relevancia la planificación de las intervenciones reparatorias a víctimas de

maltrato considerando todos los factores involucrados, ya sea maltrato individual, familiar o contextual (Paicabi, 2015). Con esto en mente es que se diseña el I-APSI.

Los resultados de este estudio dan cuenta de que el I-APSI otorga garantías psicométricas para su utilización en Chile. En particular entrega indicios de validez factorial, validez convergente y de confiabilidad (Anastasi & Urbina, 1998).

La estructura unifactorial sugiere agrupar los 10 ítems en un solo factor. Desde lo teórico se puede decir que esta estructura posee parsimonia, lo que hace recomendable su utilización en la práctica. Además, a nivel empírico esta estructura demuestra alta confiabilidad, lo que indica que las puntuaciones miden de forma consistente el apoyo institucional y que las puntuaciones obtenidas por los sujetos tienden a reflejar verdaderas diferencias en el constructo medido (Oviedo & Campo-Arias, 2005).

Por otro lado, se aprecia que el I-APSI posee validez convergente ya que su puntuación se asocia en forma directa y significativa con dos medidas de apoyo institucional que se encuentran teóricamente vinculadas (dimensiones contextuales del I-REC e I-PROT). Cabe destacar que la magnitud de estas relaciones es moderada (relaciones del orden del .5) lo que demuestra que, aunque se trata de constructos relacionados, los tres instrumentos miden formas diferentes de apoyo institucional.

Lo anterior pone en relieve la utilidad del I-APSI para medir un aspecto específico y general del apoyo institucional: la visibilización del niño y de su familia, aspecto muy necesario para el posterior reconocimiento y movilización de recursos protectores del entorno. En este estudio se pudo evidenciar cómo, en la medida que la red institucional visibiliza al niño y a su familia, también tiende a demostrar una mayor capacidad para reconocer situaciones de maltrato infantil grave y,

por ende, para actuar de forma protectora, lo que es altamente deseable en las intervenciones con víctimas de este tipo de vulneraciones contra los niños.

No obstante, por tratarse de un estudio realizado con una muestra relativamente pequeña, es necesario considerar estos resultados como preliminares. Futuros estudios deberán profundizar en los análisis, por ejemplo, mediante un análisis factorial confirmatorio o mediante otras pruebas de validez, como la validez predictiva. En este sentido, sería relevante estudiar el valor predictivo de un alto apoyo de la red social del niño en su proceso de recuperación emocional.

Por último, es necesario destacar que al tratarse de un instrumento heteroaplicado, es necesario que los profesionales que lo utilicen sean entrenados en el formato de su aplicación según las directrices establecidas en el manual (Paicabi, 2012). Se considera que aplicar el instrumento sin el entrenamiento adecuado podría llevar a tomar decisiones erradas respecto del proceso de intervención con niños, niñas y adolescentes cuyos derechos han sido gravemente vulnerados.

Referencias

- Arredondo, V., Saavedra, C., & Guerra, C. (2017). Análisis psicométrico preliminar de un instrumento para evaluar el reconocimiento de situaciones de maltrato infantil (I-REC) en Chile. *Revista de Psicología Clínica con Niños y Adolescentes*, 4(2), 111-117. Recuperado de <http://www.revistapcna.com/es>
- Arredondo-Ossandon, V., Saavedra-Inostroza, C., & Guerra-Vío, C. (2016). Análisis psicométrico preliminar de un instrumento para evaluar indicadores protectores en situaciones de maltrato infantil grave (I-PROT). *Revista de Trabajo Social*, 90, 3-14. Recuperado de <http://rts.alerta.cl/index.php/rts/index>
- Anastasi, A., & Urbina, S. (1998). *Test psicológicos*. México: Prentice Hall.
- Barudy, J. (1998). *El dolor invisible de la infancia. Una lectura ecosistémica del maltrato infantil*. Barcelona: Paidós.
- Barudy, J. (2001). *Maltrato Infantil. Ecología social: Prevención y reparación*. Santiago de Chile: Galdoc.
- Belsky, J. (1980). Child maltreatment, an ecological integration. *American Psychologist*, 35(4), 320-335. doi: 10.1037/0003-066X.35.4.320
- Bronfenbrenner, U. (1987). *La Ecología del Desarrollo Humano*. Barcelona: Paidós.
- Cicchetti, D., & Rogosch, F. A. (1997). The role of self-organization in the promotion of resilience in maltreated children. *Development and Psychopathology*, 9, 797-815. doi: 10.1017/s0954579497001442
- Cicchetti, D., Rogosch, F. A., Lynch, M., & Holt, K. D. (1993). Resilience in maltreated children: Processes leading to adaptive outcome. *Development and Psychopathology*, 5, 629-647. doi: 10.1017/s0954579400006209
- Cyrulnik, B. (2003). *Los patitos feos. La resiliencia: Una infancia infeliz no determina la vida*. Madrid: Gedisa.
- Dominguez-Lara, S. (en prensa). Fiabilidad y alfa ordinal. *Actas Urológicas Españolas*.
- Egeland, B., Carlson, E., & Sroufe, L. A. (1993). Resilience as process. *Development and Psychopathology*, 5(4), 517-528. doi: 10.1017/S0954579400006131
- Evans, S. E., Steel, A. L., & DiLillo, D. (2013). Child maltreatment severity and adult trauma symptoms: Does perceived social support play a buffering role? *Child Abuse & Neglect*, 37, 934-943. doi: 10.1016/j.chiabu.2013.03.005
- Flores, E., Cicchetti, D., & Rogosch, F. A. (2005). Predictors of resilience in maltreated and nonmaltreated latino children. *Developmental Psychology*, 41(2), 338-351. doi: 10.1037/0012-1649.41.2.338
- Gracia, E., & Musitu, G. (1993). *El maltrato infantil. Un análisis ecológico de los factores de riesgo*. Madrid: Centro de Publicaciones de Ministerio de Asuntos Sociales.

- Guerra, C., & Arredondo, V. (2017). Investigación sobre psicoterapia en abuso sexual infantil: ¿Una tarea pendiente en Chile? *Summa Psicológica* 14(1), 1-11. Recuperado de <https://summapsicologica.cl/index.php/summa>
- Guerra, C., Ocaranza, C., & Weinberger, K. (2016). Searching for social support moderate the relationship between polyvictimization and externalizing symptoms: A brief report. *Journal of Interpersonal Violence*, (Advance online publication) doi: 10.1177/0886260516642293
- Guerra, C., Pereda, N., Guilera, G., & Abad, J. (2016). Internalizing symptoms and polyvictimization in a clinical sample of adolescents: The roles of social support and non-productive coping strategies. *Child Abuse & Neglect*, 54, 57-65. doi: 10.1016/j.chiabu.2016.03.004
- Kline, R. B. (2005). *Principles and practice of structural equation modeling*. New York: Guilford.
- Luthar, S. S., Cicchetti, D., & Becker, B. (2000). The construct of resilience: A critical evaluation and guidelines for future work. *Child Development*, 71(3), 543-562. doi: 10.1111/1467-8624.00164
- Luthar, S. S., & Goldstein, A. (2004). Children's exposure to community violence: Implications for understanding risk and resilience. *Journal Clinical Child Adolescence Psychology*, 33(3), 499-505. doi: 10.1207/s15374424jccp3303_7
- Ministerio del Interior (2013). *Encuesta Nacional de Victimización por Violencia Intrafamiliar y Delitos Sexuales*. Recuperado de https://estudios.sernam.cl/documentos/?eMjM1MTAxOQ==Encuesta_Nacional_de_Victimizaci%C3%B3n_por_Violencia_Intrafamiliar_y_Delitos_Sexuales_2012
- Morelato, G. (2011). Maltrato infantil y desarrollo: Hacia una revisión de los factores de resiliencia. *Pensamiento Psicológico*, 9(17), 83-96. Recuperado de <http://revistas.javerianacali.edu.co/index.php/pensamientopsicologico>
- Oviedo, H. C., & Campo-Arias, A. (2005). Aproximación al uso del coeficiente alfa de Cronbach. *Revista Colombiana de Psiquiatría*, 34, 572-580.
- Paicabi (2012). Protocolo de aplicación de instrumento de medición de indicadores de Apoyo Social Institucional. Manuscrito inédito. ONG Paicabi, Viña del Mar, Chile.
- Paicabi (2015). Propuesta técnica ampliada programas reparación maltrato infantil grave. Manuscrito inédito. Viña del Mar, Chile: ONG Paicabi.
- Pereda-Beltran, N. (2009). Consecuencias psicológicas iniciales del abuso sexual infantil. *Papeles del Psicólogo*, 30(2), 135-144. Recuperado de <http://www.papelesdelpsicologo.es>
- Pinto-Cortez, C., & Venegas-Sanhueza, K. (2015). Experiencias de victimización y polivictimización en jóvenes chilenos. *Señales*, 9(14), 5-25. Recuperado de <http://www.sename.cl/revistas-senales>
- SENAME (2011). *Bases técnicas específicas. Programa de protección especializado. Modalidad maltrato infantil grave y abuso sexual*. Santiago de Chile: Sename.
- Terol-Cantero, M. del C., López-Roig, S., Neipp-López, M. C., Rodríguez-Marín, J., Pastor, M. A., & Martín-Aragón, M. (2004). Apoyo social e instrumentos de evaluación: Revisión y clasificación. *Anuario de Psicología*, 35(1), 23-45.
- Turner, H. A., Shattuck, A., Finkelhor, D., & Hamby, S. (2015). Effects of poly-victimization on adolescent social support, self-concept, and psychological distress. *Journal of Interpersonal Violence*, 32(5), 755-780. doi: 10.1177/0886260515586376
- UNICEF (1989). Convención de los Derechos del Niño. Recuperado el 22 de octubre del 2017 de <https://www.unicef.es/causas/derechos-ninos/convencion-derechos-ninos>
- UNICEF (2012). *4º Estudio de Maltrato Infantil en Chile. Análisis comparativo 1994-2000-2006-2012*. Santiago: UNICEF. Recuperado de <http://unicef.cl/web/wp-content/uploads/2015/07/Maltrato-Infantil.pdf>

Valentine, L., & Feinauer, L. L. (1993). Resilience factors associated with female survivors childhood sexual abuse. *The American Journal of Family Therapy*, 21, 216-224.

Vivanco, M. (1999). *Análisis estadístico multivariante: Teoría y práctica*. Santiago, Chile: Editorial Universitaria.

Adaptación de la Escala de Prosocialidad de Caprara en Adolescentes Argentinos

Adaptation of the Caprara Prosociality Scale in Argentine Adolescents

Lucas Marcelo Rodríguez *^{1,2}, Belén Mesurado¹, María Emilia Oñate^{1,2},
Paulina Guerra¹, María Soledad Menghi^{1,2}

1 - Centro Interdisciplinario de Investigaciones en Psicología Matemática y Experimental (CIIPME-CONICET), Argentina.

2 - Universidad Católica Argentina, Facultad "Teresa de Ávila", Argentina.

Recibido: 26/08/2017 Revisado: 14/09/2017 Aceptado: 01/10/2017

Introducción
Metodología
Resultados
Discusión
Referencias

Resumen

El concepto de *prosocialidad* comprende aquellos comportamientos tendientes a ayudar o beneficiar a otras personas. Es de gran importancia su medición en la población para una adecuada convivencia social. Los objetivos del presente trabajo fueron: estudiar la validez de construcción de la traducción española de la Escala de Prosocialidad de Caprara, Steca, Zelli, & Capanna (2005) en una población de adolescentes argentinos, mediante la comparación de tres modelos teóricos, y estudiar la validez convergente y los niveles de confiabilidad de la versión en español de la Escala de Prosocialidad. Se conformó una muestra de 737 adolescentes de ambos sexos, 274 varones y 463 mujeres. La edad de los mismos fluctuó entre 15 y 18 años. El modelo que obtuvo un mejor ajuste fue el modelo de 10 ítems con dos dimensiones, a saber: *conducta prosocial* por un lado, y *empatía y apoyo emocional* por otro. La escala completa obtuvo un coeficiente alfa de .78. Ambas dimensiones correlacionaron positivamente con las tendencias prosociales *sensible, anónima y pública*.

Palabras clave: *prosocialidad, empatía, adolescencia, escala, validez*

Abstract

Prosociality is understood as types of behaviour intended to help or benefit other people. Its measurement is of great importance for proper social cooperation. The aims of the present work were: to study the construct validity of the Spanish version of the Caprara, Steca, Zelli & Capanna's Prosociality Scale (2005) in Argentine adolescents, by comparing three theoretical models, and to study the convergent validity, and levels of reliability of the Spanish version of the Prosociality Scale. The sample included 737 adolescents - 274 boys and 463 girls-, between 15 and 18 years old. The best fitted model was the 10 item- model with two dimensions: *prosocial behavior*, and *empathy and emotional support*. Cronbach's alpha value relative to the whole scale was .78. Both dimensions positively correlated with the prosocial tendencies sensitive, anonymous and public.

Keywords: *prosociality, empathy, adolescence, scale, validity*

*Correspondencia a: Lucas M. Rodríguez, Buenos Aires 239 CP 3100. Paraná, Entre Ríos, Argentina. lucasmarcelorodriguez@gmail.com

Cómo citar este artículo: Rodríguez, L. M., Mesurado, B., Oñate, M. E., Guerra, P., & Menghi, M. S. (2017). Adaptación de la Escala de Prosocialidad de Caprara en adolescentes argentinos. *Revista Evaluar*, 17(2), 177-187. Recuperado de <https://revistas.unc.edu.ar/index.php/revaluar>

Introducción

El estudio de la prosocialidad ha progresado en los últimos años, dada la importancia en el transcurso del desarrollo de una personalidad sana, orientada a una relación interpersonal y social positiva. La investigación de la prosocialidad se ha incrementado en las últimas décadas (Auné & Attorresi, 2017; Caprara, 2005; Caprara & Pastorelli, 1993; Carlo, Eisenberg, & Knight, 1992; Carlo & Randall, 2002; Eisenberg & Fabes, 1998; Richaud & Mesurado, 2016; Roche-Olivar, 1997, 1999, 2010; Rodríguez, 2014).

Son considerados prosociales aquellos comportamientos tendientes a ayudar o beneficiar a otras personas, independientemente de la intención que se persiga con esta ayuda. El comportamiento prosocial es el resultado de múltiples factores individuales y situacionales que incluyen variables parentales y rasgos empáticos (Eisenberg & Fabes, 1998).

La prosocialidad es entendida como la tendencia a dar curso a acciones que se caracterizan por el efecto beneficioso que producen en el otro, y pertenece a la esfera de los hábitos, las prácticas y la modalidad habitual de la interacción social (Caprara, 2005). Hay diferentes tipos de conductas prosociales, motivadas por razones diversas, como por ejemplo la necesidad de que otro vea que la persona ayuda, la demanda externa por parte de quien es ayudado, el altruismo, etc. (Carlo & Randall, 2002). Algunos autores consideran que el altruismo consiste en la ayuda a otros, principalmente por orientación a los demás o por razones morales, sin tener en cuenta las recompensas o castigos externos (Carlo, Eisenberg, Troyer, Switzer, & Speer, 1991).

Roche-Olivar (1997, 1999, 2010) expresa que la verdadera conducta prosocial aparece al ayudar a otras personas o grupos en ausencia de recompensa extrínseca o material. Así, este autor

define la prosocialidad de la siguiente manera:

Aquellos comportamientos que, sin la búsqueda de recompensas externas, extrínsecas o materiales, favorecen a otras personas o grupos (según los criterios de éstos) o metas sociales, objetivamente positivas y aumentan la probabilidad de generar una reciprocidad positiva de calidad y solidaria en las relaciones interpersonales o sociales consecuentes, salvaguardando la identidad, creatividad e iniciativa de los individuos o grupos implicados (Roche Olivar, 1999, p. 19).

Existen diversas acciones que contribuyen al comportamiento de tipo prosocial, como: ayuda física, ayuda verbal, consuelo verbal, confirmación y valorización positiva del otro, escucha profunda, empatía, solidaridad, presencia positiva y unidad (Roche-Olivar, 1999).

Los inconvenientes o límites de la prosocialidad, según este autor, son aquellos comportamientos prosociales que no están ordenados de manera que preserven o salvaguarden la identidad, creatividad e iniciativa de las personas o grupos implicados, lo que puede generar consecuencias negativas e indeseables, como la disminución de la autoestima del receptor, los sentimientos de inferioridad, el aprendizaje de la dependencia, el aumento de la autoestima del autor a costa de un aumento del dominio sobre el receptor. De todos modos, los comportamientos prosociales, mejoran la autoestima de los autores de los mismos, su conciencia y la percepción de sus capacidades, aunque este no sea el objetivo que motive la conducta prosocial (Roche-Olivar, 1999).

La prosocialidad requiere de un desarrollo complejo que necesita cierto nivel de maduración de la persona, por lo que con la edad se producen cambios en la conducta prosocial. Durante la infancia, el egocentrismo del niño hace que sus conductas se basen en recompensas visibles. Con el paso del tiempo, las recompensas se van sustituyendo por incentivos sociales, la internalización

de normas, o el autoconcepto positivo al percibirse prosocial (Martí-Vilar & Lorente-Escriche, 2010).

En la adolescencia y la adultez, la conducta prosocial adquiere estabilidad vinculada al desarrollo sociomoral y a los procesos de razonamiento moral. Estudios longitudinales indican que, en las mujeres, la prosocialidad se mantiene estable durante la adolescencia; mientras que en los varones, aumenta entre los 15 y 19 años (Vecchione & Picconi, 2005).

A partir de la adolescencia, la capacidad de empatía alcanza una forma más madura (Bonino, 2005); de esta manera, la persona puede percibir los estados experimentados por otros individuos en situaciones en que la ayuda puede ser más o menos solicitada de forma expresa. De este modo, la implicación afectiva a partir de la adolescencia, junto con una plena conciencia de las intenciones y efectos de sus acciones, determina un desarrollo maduro de la persona prosocial (Paciello, Vecchio, & Pepe, 2005). Esta argumentación es el fundamento para incorporar la empatía como dimensión integrante de la prosocialidad en la adultez (Caprara, Capanna, Steca, & Paciello, 2005; Caprara, Steca, Zelli, & Capanna, 2005) y en la adolescencia (Paciello et al., 2005). Este giro conceptual genera el debate sobre la empatía como predictor o como dimensión integrante de la prosocialidad (Auné, Abal, & Attorresi, 2015a).

Hay resultados contradictorios en cuanto a las diferencias en la prosocialidad según el sexo de los sujetos (Martí-Vilar & Lorente-Escriche, 2010; Vecchione & Picconi, 2005). Esta falta de acuerdo podría deberse a los diversos instrumentos de medición de la variable prosocialidad, que podrían focalizarse en distintos aspectos (Vecchione & Picconi, 2005). Otra causa posible en relación a las diferencias detectadas en la prosocialidad es que el tipo de comportamiento prosocial varía según el sexo. Según Auné, Abal, & Attorresi (2015b), no se observan diferencias de

género en conductas prosociales que requieren esfuerzo alto o muy poco esfuerzo por parte de quien las practica. Sin embargo, los autores encontraron que las mujeres presentan más conductas prosociales con exigencias bajas-medias, empáticas y de cuidado y entendimiento emocional.

Como la prosocialidad implica un rango amplio de conductas, al medirla se puede realizar una distinción entre medidas de conducta prosocial global y conductas prosociales medidas en situaciones específicas (Carlo & Randall, 2002). Las medidas de conducta prosocial global son las más comunes y se definen como medidas que evalúan las tendencias personales de exhibición de una serie de comportamientos prosociales en todos los contextos y motivos, y un subtipo de las mismas son las medidas de competencias sociales. Las medidas de conducta prosocial global tienden a caracterizar a las personas prosociales, distinguiéndolas de aquellas que no lo son. Estas medidas podrían limitar la tarea investigativa en cuanto al abordaje de cuestiones conceptuales específicas respecto a la correlación de los comportamientos prosociales, ya que no distinguen si la prosocialidad es espontánea o ayudada por un requerimiento externo, si hay una motivación intrínseca o extrínseca, etc. Como ejemplo de medidas globales puede citarse la Escala de Prosocialidad de Caprara, Steca, et al. (2005), con la que recientemente se han realizado estudios en poblaciones de adolescentes y adultos jóvenes argentinos (Regner, 2009; Regner & Vignale, 2008; Rodríguez, 2014, 2016; Rodríguez & Moreno, 2016).

Por su parte, las medidas de conductas prosociales en situaciones específicas pueden brindar mayor información acerca de conceptualizaciones más delimitadas, así como del desarrollo y correlación de los diversos tipos de conducta prosocial, ya que conciben a la prosocialidad como una variable multidimensional, un abordaje muy beneficioso cuando se estudia su relación con otras

variables (Carlo & Randall, 2002). La Escala de Tendencias Prosociales de Carlo & Randall (2002) es un ejemplo de medida de conducta prosocial en situaciones específicas, y evalúa las diversas tendencias o motivaciones para la acción prosocial. Esta escala ha sido aplicada en estudios de poblaciones de adolescentes argentinos en relación con los estilos parentales (Richaud, Mesurado, & Lemos, 2012), y también en la investigación de las variables motivadoras de la concreción de la conducta prosocial (Rodríguez, 2014). Otro ejemplo de medidas de conducta prosocial en situaciones específicas es el caso de las escalas que permiten discriminar al receptor de la ayuda, midiendo la prosocialidad dirigida hacia familiares, amigos y extraños en población adolescente (Padilla-Walker & Christensen, 2011; Padilla-Walker, Dyer, Yorgason, Fraser, & Coyne, 2015). También se han realizado estudios con esta distinción en una población de adultos jóvenes argentinos (Mesurado & Richaud, 2017).

Si bien hay instrumentos que miden la prosocialidad en español, los instrumentos destinados a niños y adolescentes son más numerosos y de probadas propiedades psicométricas (Auné & Attorresi, 2017), por lo que es necesario contar con instrumentos que midan este constructo en poblaciones adultas. En el caso de la Escala de Prosocialidad de Caprara, Steca, et al. (2005), este instrumento mide el constructo en adolescentes, jóvenes y adultos, lo que constituiría un aporte a la necesidad planteada.

Los objetivos del presente trabajo son:

1. Estudiar la validez de constructo de la traducción española de la Escala de Prosocialidad de Caprara, Steca, et al. (2005) en una población de adolescentes argentinos, mediante la comparación de tres modelos teóricos.

2. Estudiar la validez convergente y los niveles de confiabilidad de la versión en español de la Escala de Prosocialidad.

Metodología

Muestra

Participaron en el estudio 737 adolescentes de ambos sexos, 274 varones (37.2%) y 463 mujeres (62.8%). La edad de los participantes fluctuó entre 15 y 18 años, siendo la media 16.72 (DE = .72). Los adolescentes pertenecían a escuelas de nivel medio de las provincias de Entre Ríos y de Buenos Aires.

Instrumentos

Se utilizó la Escala de Prosocialidad (EP; Caprara, Steca, et al., 2005) en su versión traducida al español (población argentina) de Regner y Vignale (2008). Dicha escala permite medir conductas prosociales en adolescentes y adultos mediante un puntaje total, y permite discriminar sujetos mayormente prosociales de los que no lo son. La escala consta de 16 ítems (Anexo 1) para responder mediante una escala tipo Likert de 5 opciones, que van desde *nunca/casi nunca a siempre/casi siempre*. A partir de los estudios realizados por los creadores del instrumento, pueden distinguirse los ítems que dan mayor información (ítems 3, 5, 7, 8, 10, 12 y 13), los ítems que dan mediana información (ítems 4, 6 y 9) y los ítems que proporcionan menor información (ítems 1, 2, 11, 14, 15 y 16) a fin de distinguir entre sujetos mayormente prosociales y aquellos que no lo son, según las propiedades psicométricas de la versión original (Caprara, Steca, et al., 2005; Paciello et al., 2005). La escala incluye ítems que comprenden ayuda, como por ejemplo: *Intento ayudar a los demás*, así como ítems que comprenden empatía, como: *Comparto intensamente las emociones de los demás*.

Con relación a la consistencia interna del instrumento, los autores originales informaron un

alfa de Cronbach de .91 (Caprara et al., 2005), mientras que investigaciones recientes en adolescentes de Argentina han informado un alfa de Cronbach de .84 (Rodríguez, 2014).

Además, se utilizó la Escala de Tendencias Prosociales (Prosocial Tendencies Measure, PTM; Carlo & Randall, 2002), adaptada al español por Richaud, Mesurado y Kohan-Cortada (2012). La escala PTM original consta de 21 ítems, que deben ser respondidos en una escala tipo Lickert de 5 opciones, que van desde *No me describe en absoluto* a *Me describe muy bien*; y evalúa seis aspectos de la tendencia prosocial: público, altruista, anónimo, emocional, crisis y complaciente. Sin embargo, la adaptación argentina presenta solamente cuatro dimensiones: público, altruista, anónimo y sensible. Dentro de la dimensión sensible se agruparon las tendencias: emocional, crisis y complaciente descritas en la escala original de Carlo y Randall (2002).

Con relación a la consistencia interna del instrumento, en la versión en español se informó un alfa de Cronbach para toda la escala de .71 en adolescentes.

Procedimiento

Para realizar la recolección de datos se contactaron instituciones educativas de nivel medio de las provincias de Entre Ríos y Buenos Aires. Luego de obtener el permiso de las autoridades de las instituciones, se dialogó con los alumnos para explicarles su participación en la investigación, la cual era de carácter anónimo. Se le solicitó a cada participante que firmara el consentimiento informado, si tenían al menos 18 años, y si eran menores de 18 años, que lo firmaran sus padres o tutor.

Análisis de datos

A fin de evaluar la validez de constructo del instrumento se utilizó el análisis factorial confirmatorio (AFC), mediante el método de estimación de parámetros por máxima verosimilitud, con el programa AMOS 17 (IBM, 2009). Para evaluar el ajuste del modelo se consideraron los siguientes índices: chi cuadrada (χ^2), chi cuadrada dividida por los grados de libertad (χ^2/gl), índice comparativo de ajuste (CFI), índice de bondad de ajuste (GFI) e índice ajustado de bondad de ajuste (AGFI). Como medidas del error se utilizaron la raíz de cuadrados medios del error (RMSR) y el promedio de los residuales estandarizados al cuadrado (RMSEA). A fin de obtener un buen ajuste de los datos al modelo teórico propuesto por AFC, el valor de los índices: índice comparativo de ajuste (CFI), índice de bondad de ajuste (GFI) e índice ajustado de bondad de ajuste (AGFI) debe ser superior a .90. Asimismo, el valor de χ^2/gl debe ser inferior a 2, y las medidas del error RMSR y RMSEA deben tener un valor igual o menor de .60 para un ajuste óptimo (Kline, 1998; Cupani, 2012).

La validez convergente se evaluó mediante correlaciones de Pearson; mientras que la confiabilidad del instrumento, mediante el índice alfa de Cronbach.

Resultados

Para evaluar la validez de constructo se llevaron a cabo tres análisis factoriales confirmatorios (AFC) por medio del método de estimación de parámetros por máxima verosimilitud. En el primer análisis, denominado Modelo 1, se incluyeron los 16 ítems propuestos por Caprara, Steca, et al. (2005) en una sola dimensión (versión original). Sin embargo, dado que este modelo no obtu-

Tabla 1

Ajuste de los modelos de medición de la Escala de Prosocialidad evaluados mediante AFC.

Modelo	χ^2	gl	p	χ^2/gl	GFI	AGFI	CFI	RMR	RMSEA
Modelo 1	636.81	104	.001	6.12	.89	.86	.77	.071	.083
Modelo 2	530.33	103	.001	5.15	.92	.89	.81	.065	.075
Modelo 3	192.93	34	.001	5.67	.95	.93	.90	.054	.08

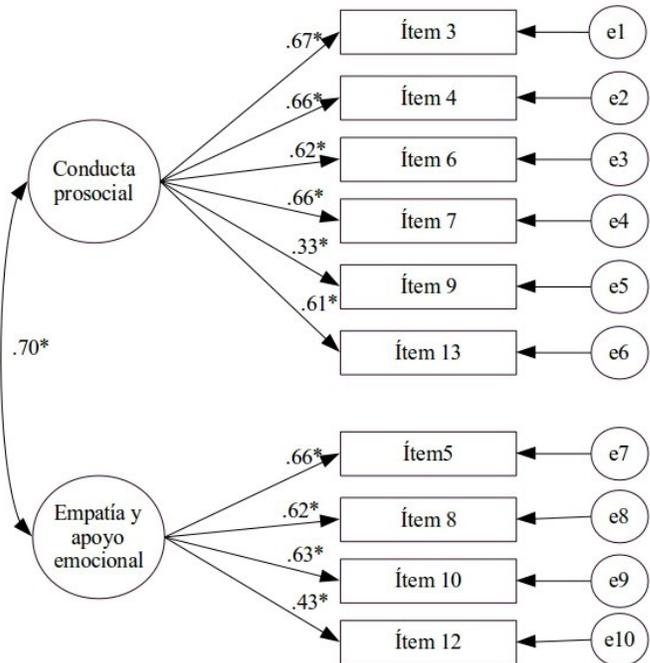
vo buen ajuste (Tabla 1, Modelo 1), se plantearon dos modelos alternativos.

El Modelo 2 se realizó con los 16 ítems propuestos, pero divididos en dos dimensiones según lo propuesto en la teoría: una dimensión que evalúa empatía y apoyo emocional (ítems 5, 8, 10 y 12), y otra que evalúa propiamente la conducta prosocial (ítems restantes). Por último, se probó un tercer modelo (Modelo 3), que conserva las dos dimensiones citadas pero incluye solamente 10 ítems. Los ítems 1, 2, 11, 14, 15 y 16 fueron omitidos por considerar que brindan menor información, según análisis psicométricos previos realizados por los autores originales de la Escala de Prosocialidad (Caprara, Steca, et al., 2005; Paciello et al., 2005).

La Tabla 1 muestra los ajustes de cada uno de los modelos probados mediante AFC. A partir de la comparación de los tres modelos determinamos que el Modelo 3 obtuvo el mejor ajuste. En la Figura 1 puede apreciarse el Modelo 3 con las cargas factoriales estandarizadas.

Para evaluar la confiabilidad de las dimensiones del Modelo 3 se determinó el índice alfa de Cronbach. El valor de alfa de la dimensión conducta prosocial fue de .71; mientras que el de la dimensión empatía y apoyo emocional fue de .67. La escala completa (10 ítems) presentó un valor alfa de .78.

Para evaluar la validez convergente de la Escala de Prosocialidad se consideró el coeficiente de correlación de Pearson, como se muestra en la Tabla 2.

**Figura 1**

Modelo 3 (2 dimensiones con 10 ítems).

Nota. * $p \leq .001$

Discusión

Para evaluar la validez de constructo de la Escala de Prosocialidad de Caprara, Steca, et al. (2005) se realizó el análisis factorial confirmatorio (AFC) de tres modelos. Teniendo en cuenta los criterios mencionados en el apartado de metodología, el Modelo 3 obtuvo el mejor ajuste.

En el Modelo 3 se consideraron dos dimensiones: *conducta prosocial* y *empatía y apoyo emocional*. En la escala original se utilizan conjuntamente bajo el presupuesto teórico de que a

Tabla 2

Correlaciones de Pearson entre las dos dimensiones de la Escala de Conducta Prosocial de Caprara y la Escala de Tendencias Prosociales.

	Conducta Prosocial (Caprara)	Empatía y Apoyo Emocional (Caprara)	TP Pública	TP Altruista	TP Anónima	TP Sensible
Conducta Prosocial (Caprara)	1					
Empatía y apoyo emocional (Caprara)	.509**	1				
TP Pública	.108**	.084*	1			
TP Altruista	.008	-.013	-.445**	1		
TP Anónimo	.308**	.115**	.016	-.003	1	
TP Sensible	.611**	.551**	.200**	-.151**	.350**	1

Nota. ** $p < .01$, * $p < .05$, $N = 737$.

partir de la adolescencia, y en adelante en el ciclo vital, la empatía es parte integrante de la prosocialidad (Caprara, Steca, et al., 2005; Paciello et al., 2005). Consideramos que si bien la empatía es una parte esencial para la concreción de la conducta prosocial, empatía y conducta prosocial son dimensiones claramente diferentes, aunque asociadas; lo que puede observarse en las correlaciones obtenidas en este estudio. Esta diferencia queda clara al analizar la formulación de la Escala de Conducta Prosocial. Por ejemplo, el ítem 8: *Comparto intensamente las emociones de los demás* (empatía); hace clara alusión a la empatía, que puede definirse como compartir la emoción que se percibe del otro, sentir con el otro (Eisenberg & Strayer, 1992) y se distingue de ítems como *Intento ayudar a los demás* (conducta prosocial), que expresa con claridad la intención de activar o realizar una conducta dirigida a ayudar a los demás. Esta diferencia en los constructos se evidenció en la prueba de los modelos mediante AFC, en la que el modelo que obtuvo un mejor ajuste fue aquel que presentó *empatía y conducta prosocial* como dimensiones independientes.

En cuanto a la consistencia interna del instrumento, los índices alfa de Cronbach mostraron una buena consistencia interna, según el número de ítems de cada dimensión, y en la totalidad de la escala en el Modelo 3, de acuerdo con los criterios de Kaplan y Saccuzzo (2006; véase también Loewenthal, 2001).

Para evaluar la validez convergente se obtuvieron correlaciones de Pearson con la Escala de Tendencias Prosociales de Carlo y Randall (2002). En cuanto a las dimensiones estudiadas, se encontró correlación positiva en los pares *sensible*, *anónimo* y *pública*, a saber: conducta prosocial y sensible, empatía y sensible, conducta prosocial y anónimo, empatía y anónimo, conducta prosocial y público, lo que implica una prueba de validez convergente. Por otro lado, las dimensiones de la Escala de Prosocialidad y la tendencia prosocial altruista no manifestaron correlación según el coeficiente de Pearson.

El hecho de que los mayores valores de correlación del instrumento se hayan presentado en relación con la tendencia prosocial *sensible* es coincidente con los hallazgos de otros estu-

dios (Carlo & Randall, 2002; Rodríguez, 2014). La tendencia prosocial sensible implica una demanda externa intensa de la persona beneficiada con la conducta prosocial (Richaud et al., 2012); comprende ayuda a los demás en respuesta a pedidos verbales o no verbales por parte del receptor de la ayuda (tendencia prosocial complaciente). Además, la conducta prosocial ayuda a otros que se encuentran en circunstancias emocionalmente evocadoras, y se asocia fuertemente con la compasión y otras tendencias de orientación hacia el otro (tendencia prosocial emocional), y con la ayuda a personas en crisis o en situación de emergencia (tendencia prosocial en crisis; Carlo & Randall, 2002). La intensa demanda externa de la persona ayudada requiere de la empatía para que sea evocada, lo que explica la alta correlación de la conducta prosocial con la empatía. Por otra parte, esta demanda externa intensa estaría vinculada al surgimiento de la conducta de ayuda propiamente dicha.

La tendencia prosocial de tipo anónima es la tendencia a mostrar conductas de ayuda con la intención de que el beneficiario de la ayuda no conozca la identidad de su benefactor (Carlo & Randall, 2002). Este tipo de tendencia prosocial mostró también una asociación con las dimensiones de *empatía* y *apoyo emocional* así como con la *conducta prosocial*. Nuestros resultados también confirman que diferentes aspectos de la intencionalidad de la conducta prosocial están acompañados de una carga empática importante hacia el beneficiario y de la conducta de ayuda.

Por su parte, la tendencia prosocial *pública* consiste en una ayuda realizada en la presencia de audiencia, con el deseo de obtener, en parte, aprobación o respeto de los demás, y mejorar la autoestima de la propia persona (Carlo & Randall, 2002). La dimensión *pública* podría tener una correlación mayor con la *conducta prosocial*, debido a que la conducta de ayuda se manifiesta de

todos modos, independientemente de los motivos por los cuales se realiza, lo que no se observa. No es así en el caso de la empatía, dimensión que no presentó correlación significativa con conducta prosocial. Puede hipotetizarse que en las tendencias prosociales de tipo *públicas* (que se realizan en busca de un beneficio personal más que de la otra persona) es lógico que esté ausente toda experiencia de empatía emocional.

Por último, no se encontró asociación entre las dimensiones *conducta prosocial* y *empatía* y *apoyo emocional* y las tendencias *altruistas*, lo que coincide con resultados obtenidos durante la validación de la versión original de PTM de Carlo y Randall (2002). Una explicación para estos resultados puede ser que la dimensión *altruista* del PTM está operacionalizada con la totalidad de sus ítems inversos (p. ej., *Creo que donar dinero es bueno cuando consigo una ventaja*), hecho que podría afectar la correcta evaluación del altruismo, dado que una respuesta negativa a este ítem no implicaría necesariamente que siempre se busque donar dinero desinteresadamente.

Cabe aclarar que el estudio de la validez convergente de la versión española de la Escala de Prosocialidad de Caprara, Steca, et al. (2005) se realizó con un instrumento que mide intencionalidad de las tendencias prosociales en los jóvenes; es decir, aquello que fundamenta la conducta prosocial. Pero la conducta prosocial es la concreción de esa conducta de ayuda, lo cual no es exactamente lo mismo (conducta y motivación). En futuros estudios sería necesario incluir otras medidas de la prosocialidad con la intención de contrastar los resultados del presente estudio sobre la validez convergente. Asimismo, sería importante validar la escala en otras poblaciones, como ser jóvenes y adultos, con el modelo original existente.

Referencias

- Auné, S. E., Abal, F. J. P., & Attorresi, H. F. (2015a). Antagonismos entre concepciones de empatía y su relación con la conducta prosocial. *Revista de Psicología, 17*(2), 137-149. doi: 10.18050/revpsi.v17n2a7.2015
- Auné, S. E.; Abal, F. J. P.; & Attorresi, H. F. (2015b). Conducta prosocial y estereotipos de género. *Praxis. Revista de Psicología, 27*, 7-19. Recuperado de <http://www.praxis.udp.cl>
- Auné, S. E., & Attorresi, H. F. (2017). Dimensionalidad de un test de conducta prosocial. *Revista Evaluar, 17*(1), 29-37. Recuperado de <https://revistas.unc.edu.ar/index.php/revaluar>
- Bonino, S. (2005). Contagio, empatía e comportamiento prosociale. En G. V. Caprara & S. Bonino (Eds.), *Il comportamento prosociale. Aspetti individuali, familiari e sociali* (pp. 23-44). Trento: Erikson.
- Caprara, G. (2005). Comportamento prosociale e prosocialità. En G. V. Caprara & S. Bonino (Eds.), *Il comportamento prosociale. Aspetti individuali, familiari e sociali* (pp. 7-22). Trento: Erikson.
- Caprara, G. V., Capanna, C., Steca, P., & Paciello, M. (2005). Misura e determinanti personali della prosocialità. Un approccio sociale cognitivo. *Giornale Italiano di Psicologia, 32*(2), 287-308. Recuperado de <https://www.rivisteweb.it/doi/10.1421/20313>
- Caprara, G., & Pastorelli, C. (1993). Early emotional instability, prosocial behaviour, and aggression: Some methodological aspects. *European Journal of Personality, 7*(1), 19-36. doi: 10.1002/per.2410070103
- Caprara, G., Steca, P., Zelli, A., & Capanna, C. (2005). A new scale for measuring adults' prosocialness. *European Journal of Psychological Assessment, 21*(2), 77-89. doi: 10.1027/1015-5759.21.2.77
- Carlo, G., & Randall, B. (2002). The development of a measure of prosocial behaviors for late adolescents. *Journal of Youth and Adolescence, 31*, 31-44.
- Carlo, G., Eisenberg, N., & Knight, G. (1992). An objective measure of adolescent's prosocial moral reasoning. *Journal of Research on Adolescence, 2*, 331-349.
- Carlo, G., Eisenberg, N., Troyer, D., Switzer, G., & Speer, A. L. (1991). The altruistic personality: In what contexts is it apparent? *Journal of Personality and Social Psychology, 61*(3), 450-458. doi: 10.1037/0022-3514.61.3.450
- Cupani, M. (2012). Análisis de ecuaciones estructurales: Conceptos, etapas de desarrollo y un modelo de aplicación. *Revista Tesis, 1*, 186-199. Recuperado de <https://revistas.unc.edu.ar/index.php/tesis>
- Eisenberg, N., & Fabes, R. (1998). Prosocial development. En W. Damon (Ed. serie) & N. Eisenberg (Ed. vol.), *Handbook of Child Psychology, Vol. 3: Social, Emotional, and Personality Development* (5ª ed., pp. 701-778). New York: John Wiley.
- Eisenberg, N., & Strayer, J. (Eds.). (1992). *La empatía y su desarrollo*. Bilbao: DDB.
- IBM (2009). *Amos* (Versión 17.0) [software de cómputo]. Chicago: IBM SPSS.
- Kaplan, R. M., & Saccuzzo, D. P. (2006). *Pruebas psicológicas: Principios, aplicaciones y temas* (6ª ed.). Madrid: International Thomson.
- Kline, R. B. (1998). *Principles and practice of structural equation modeling*. New York, NY: Guilford Press.
- Loewenthal, K. M. (2001). *An introduction to psychological tests and scales* (2ª ed.). Londres: Psychology Press.
- Martí-Vilar, M., & Lorente-Escriche, S. (2010). Factores determinantes de las conductas prosociales. En M. Martí-Vilar (Ed.), *Razonamiento moral y prosocialidad. Fundamentos* (pp. 149-168). Madrid: Editorial CCS.
- Mesurado, B., & Richaud, M. C. (2017). The relationship between parental variables, empathy and prosocial-flow with prosocial behavior toward strangers, friends, and family. *Journal of Happiness Studies, 18*(3), 843-860. doi: 10.1007/s10902-016-9748-7
- Paciello, M., Vecchio, G. M., & Pepe, S. (2005). La misura della prosocialità. En G. V. Caprara & S. Bonino (Eds.), *Il comportamento prosociale. Aspetti individuali, familiari e sociali* (pp. 45-57). Trento: Erikson.
- Padilla-Walker, L. M., & Christensen, K. J. (2011). Empathy and self-regulation as mediators between

- parenting and adolescents' prosocial behavior toward strangers, friends, and family. *Journal of Research on Adolescence*, 21(3), 545-551. doi: 10.1111/j.1532-7795.2010.00695.x
- Padilla-Walker, L. M., Dyer, W. J., Yorgason, J. B., Fraser, A. M., & Coyne, S. M. (2015). Adolescents' prosocial behavior toward family, friends, and strangers: A person-centered approach. *Journal of Research on Adolescence*, 25(1), 135-150. doi: 10.1111/jora.12102
- Regner, E. (2009). Compasión y gratitud, emociones empáticas que elicitan las conductas prosociales. En M. C. Richaud & J. E. Moreno (Eds.), *Investigación en Ciencias del Comportamiento. Avances Iberoamericanos*. Buenos Aires: Ediciones CIIPME-CONICET.
- Regner, E., & Vignale, P. (2008). Adaptación de la Escala de Conductas Prosociales de Caprara y Pastorelli. Manuscrito no publicado.
- Richaud, M. C., & Mesurado, B. (2016). Las emociones positivas y la empatía como promotores de las conductas prosociales e inhibidores de las conductas agresivas. *Acción Psicológica*, 13(2), 31-42. doi: 10.5944/ap.13.2.17808
- Richaud, M., Mesurado, B., & Kohan-Cortada, A. (2012). Analysis of dimensions of prosocial behavior in an Argentinean sample of children. *Psychological Reports: Mental & Physical Health*, 111(3), 1-10. doi: 10.2466/10.11.17.pr0.111.6.687-696
- Richaud, M., Mesurado, B., & Lemos, V. (2012). Links between perception of parental actions and prosocial behavior in early adolescence. *Journal of Child and Family Studies*, 22(5), 637-646. doi: 10.1007/s10826-012-9617-x
- Roche-Olivar, R. (Ed.) (2010). *Prosocialidad, nuevos desafíos*. Buenos Aires: Ciudad Nueva.
- Roche-Olivar, R. (1999). *Desarrollo de la inteligencia emocional y social desde los valores y actitudes prosociales en la escuela*. Buenos Aires: Ciudad Nueva.
- Roche-Olivar, R. (1997). *Psicología y educación para la prosocialidad*. Buenos Aires: Ciudad Nueva.
- Rodriguez, L. M. (2014). Motivaciones y conductas prosociales en adolescentes argentinos. *Praxis. Revista de Psicología*, 25, 69-87.
- Rodriguez, L. M. (2016). *Psicología del desarrollo moral en la adolescencia. Un modelo integrativo*. (Tesis doctoral). Buenos Aires, Facultad de Psicología y Psicopedagogía de la Pontificia Universidad Católica Argentina.
- Rodriguez, L. M., & Moreno, J. E. (2016). Posturas éticas y empatía, predictores de prosocialidad y de penalización de faltas y delitos. *revista acción psicológica*, 13(2), 43-56. doi: 10.5944/ap.13.2.17809
- Vecchione, M., & Picconi, L. (2005). Differenze di età e di genere nella condotta prosociale. En G. V. Caprara & S. Bonino (Eds.), *Il comportamento prosociale. Aspetti individuali, familiari e sociali* (pp. 59-75). Trento: Erikson.

Anexo 1
Escala original de Prosocialidad de Caprara, Steca, et al. (2005).

Ítems	Nunca/ Casi nunca	Pocas Veces	A Veces	Muchas veces	Casi siempre/ Siempre
1. Con mucho gusto ayudo a los demás en el desarrollo de sus actividades.					
2. Comparto con mis amigos y conocidos las cosas que me gustan.					
3. Intento ayudar a los demás.					
4. Estoy dispuesto a realizar actividades de voluntariado en favor de los necesitados.					
5. Me conecto con el estado de ánimo del que sufre.					
6. Acudo inmediatamente en auxilio de quien lo necesita.					
7. Enseguida me pongo manos a la obra cuando otros se encuentran en dificultades.					
8. Comparto intensamente las emociones de los demás.					
9. Con mucho gusto pongo a disposición de los demás mis conocimientos y capacidades.					
10. Intento consolar a quien está triste.					
11. Presto espontáneamente dinero u otras cosas.					
12. Me resulta fácil ponerme en el lugar de quien está disgustado.					
13. Trato de estar cerca y cuidar de quien lo necesita.					
14. Comparto con placer con mis seres queridos o conocidos las buenas oportunidades que se me ofrecen.					
15. Con mucho gusto hago compañía a los conocidos que se sienten solos.					
16. Intuyo el malestar de los demás aún cuando no me lo comuniquen directamente.					

Nota. La escala validada en este estudio comprende 10 ítems, los que se calcularon del siguiente modo: la dimensión conducta prosocial se obtuvo del promedio de 6 ítems (3, 4, 6, 7, 9 y 13); mientras que la dimensión empatía y apoyo emocional se obtuvo del promedio de 4 ítems (5, 8, 10, 12).