

Aplicación del Modelo de Escala de Clasificación Para Examinar las Propiedades Psicométricas de la Self-Control Scale

Application of the Rating Scale Model to Examine the Psychometric Properties of the Self-Control Scale

Sebastian Jesús Garrido, Marcos Cupani y Karin Arbach
Universidad Nacional de Córdoba

Centro de Investigación de la Facultad de Psicología (CIPsi) - Grupo Vinculado
Centro de Investigaciones y Estudios sobre Cultura y Sociedad (CIECS) - CONICET

Considerando la importancia teórica, empírica y práctica de la valoración del autocontrol mediante instrumentos válidos y confiables, en el presente estudio se evaluaron las propiedades psicométricas de la versión original de la *Self-Control Scale* (SCS). Con este fin, se utilizó el modelo de escala de clasificación para estimar la unidimensionalidad, la idoneidad de las categorías de respuesta y el ajuste de las personas a los ítems de la SCS. Estos análisis se realizaron en una muestra por conveniencia de 1072 estudiantes universitarios argentinos. Entre los principales resultados puede mencionarse que la SCS permite una valoración unidimensional del constructo autocontrol y que una versión de la escala con 3 categorías de respuesta podría ser más adecuada para la medición del autocontrol que la versión original de 5 categorías. Sin embargo, se encontró que con la nueva configuración de las categorías de respuesta se presentaba un funcionamiento diferencial en 3 ítems de acuerdo al sexo de los participantes. Por último, el mapa de distribución de ítems y de personas mostró ausencia de elementos en el extremo superior e inferior del continuo autocontrol.

Palabras clave: autocontrol, modelo Rasch, Self-Control Scale, propiedades psicométricas, estudiantes universitarios

Considering the theoretical, empirical, and practical importance of self-control assessment through valid and reliable instruments, the present study examined the psychometric properties of the original version of the *Self-Control Scale* (SCS). To this end, the rating scale model was used to estimate unidimensionality, the adequacy of response categories, and participant fit. These analyses were conducted using a convenience sample of 1072 Argentine university students. Among other findings, it was observed that the SCS makes it possible to assess the self-control construct unidimensionally and that a version of the scale with 3 response categories would be more adequate for measuring self-control than the original 5-category version. However, it was found that the new configuration of response categories led to differential functioning in 3 items depending on the participants' sex. Finally, the person-item distribution map showed a lack of elements at the top and bottom of the self-control continuum.

Keywords: self-control, Rasch model, Self-Control Scale, psychometric properties, university students

Se ha definido al constructo *autocontrol* como la capacidad que posee el individuo para controlar las acciones, pensamientos o emociones impulsivas o no planificadas que, generalmente, se encuentran dirigidas a la obtención de una gratificación inmediata (Dixon & Tibbetts, 2009). En la literatura actual este constructo se considera una de las estructuras de personalidad más relevantes, debido a la incidencia que posee en la capacidad de adaptación del individuo, según presenten bajos o altos niveles de autocontrol (Vohs, Finkenauer & Baumeister, 2011). Efectivamente, altos niveles de autocontrol se han asociado con el compromiso en objetivos a largo plazo, mayor inhibición de la respuesta emocional negativa y un mejor ajuste social, debido a la predisposición para modificar y adecuar su conducta a las normas sociales (de Ridder, Lensvelt-Mulders, Finkenauer, Stok & Baumeister, 2012). Por el contrario, bajos niveles de autocontrol se asocian con comportamiento de riesgo, como el consumo de drogas, conflictos en las relaciones personales y dificultad para permanecer en instituciones que establezcan disciplina, supervisión o límites (Moon, McCluskey, Blurton & Hwang, 2014).

Sebastian Jesús Garrido, Marcos Cupani y Karin Arbach, Facultad de Psicología, Universidad Nacional de Córdoba, Centro de Investigación de la Facultad de Psicología (CIPsi), Grupo Vinculado Centro de Investigaciones y Estudios sobre Cultura y Sociedad (CIECS) del Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina.

La correspondencia relativa a este artículo debe ser dirigida a Sebastian Jesús Garrido, CIPsi, Grupo Vinculado Centro de Investigaciones y Estudios sobre Cultura y Sociedad (CIECS)-CONICET, Facultad de Psicología, Universidad Nacional de Córdoba, Enfermera Gordillo esquina Enrique Barros, Ciudad Universitaria 5000, Córdoba, Argentina. E-mail: sebagarrido@gmail.com

Considerando la relevancia del constructo autocontrol, se elaboraron diversas medidas para evaluarlo (de Ridder et al., 2012; Duckworth & Kern, 2011). En efecto, hasta el año 2011 se identificaron más de 100 autoinformes para la medición del autocontrol (Duckworth & Kern, 2011), entre los que se destacan por su utilización la *Barratt Impulsiveness Scale* (Patton, Stanford & Barrat, 1995), la *Low-Self-Control Scale* (Grasmick, Tittle, Bursik & Arneklev, 1993) y la *Self-Control Scale* (SCS; Tangney, Baumeister & Boone, 2004). Con respecto a esta última escala, puede decirse que su amplia utilización, tal vez, responda a su capacidad para valorar la relación entre el autocontrol y distintos dominios del comportamiento (de Ridder et al., 2012).

Un volumen extenso de investigaciones (Ludwig et al., 2013; Malouf, Stuewig & Tangney, 2012; Prado, Treeby & Crowe, 2015) utilizaron la versión abreviada de la SCS, la cual posee altos niveles de correlación con la versión extensa de la escala y contempla el mismo contenido que la escala total de autocontrol (Tangney et al., 2004). Si bien en su gran mayoría estos estudios reportaron adecuados índices de confiabilidad, la estructura factorial original del instrumento fue difícil de replicar (Maloney, Grawitch & Barber, 2012; Morean et al., 2014), lo que generó que algunos autores consideren una estructura factorial bidimensional como la más adecuada (Lindner, Nagy & Retelsdorf, 2015), aunque los autores originales de la SCS consideren que es unidimensional (Tangney et al., 2004).

Debido a los modelos teóricos contradictorios que derivaron de los diferentes análisis factoriales reportados en la literatura, se necesita una evaluación psicométrica más rigurosa para evaluar la validez de constructo de la escala, mediante modelos teóricos superadores (Jones & Thissen, 2007). Como regla general, la teoría psicométrica que se empleó para la construcción del test es la teoría clásica de los test (TCT). Las características métricas de esta teoría, sin embargo, presentan un problema doble. En primer lugar, las mediciones para cada participante dependen del instrumento que se haya utilizado. En segundo lugar, las estimaciones de los ítems y las propiedades de los test dependen de la muestra de individuos utilizada, por ejemplo, la confiabilidad del test dependerá de la muestra de participantes empleada para calcularla. Además, la TCT supone que una vez que la confiabilidad de una prueba ha sido estimada para una cierta población, esta precisión se mantiene constante para todos los niveles (por ejemplo, asume que la confiabilidad seguirá siendo idéntica al estimar las medidas de las personas con valores bajos, medios y altos en el rasgo).

Los avances en el campo de la psicometría ayudaron a trasladarnos desde la TCT a la teoría de respuesta a los ítems (TRI). Con la TRI se pueden obtener medidas constantes, es decir, medidas que son independientes de los instrumentos utilizados y de los individuos evaluados (Hambleton & Swaminathan, 1985). Incluso, el proceso de calibración (Embretson & Reise, 2000) es independiente de la muestra utilizada para calibrar el test y la medición de las personas también es independiente del test (no importa qué ítems se utilizaron para evaluar a los participantes). Además, pierde sentido referirse a la confiabilidad total del test desde la TRI, porque desde esta teoría la precisión de medida se estima para cada nivel de habilidad o rasgo. Por otra parte, los modelos de la TRI se enfocan en la medición conjunta de las personas y de los ítems, los cuales se colocan en una misma métrica de medida en el continuo, a diferencia de la TCT, que solo localiza a las personas en el continuo y asume que todos los ítems contribuyen de la misma forma a la medición del constructo.

En efecto, en la actualidad la TRI y el modelo de Rasch en particular, están ganando popularidad por encima de la TCT como modelos para crear y validar instrumentos (Hambleton, 2000). Esto se explica porque el modelo de Rasch proporciona una metodología completa y detallada que permite evaluar las propiedades psicométricas de un instrumento a nivel de los ítems (Messick, 1994). El análisis de Rasch proporciona más información acerca de la capacidad de una persona, porque se centra en la dificultad de los ítems más que en la cantidad de ítems que responde correctamente cada participante. Desde la perspectiva de Rasch, la habilidad de una persona interactúa con la dificultad del ítem para poder así obtener una puntuación para cada sujeto en la medida (Linacre, 2002). Sin embargo, no todos los modelos de TRI son adecuados para cuantificar las mismas variables. Por el contrario, se debe seleccionar el modelo según la naturaleza de la variable que se pretende cuantificar, como así también la forma en que las pruebas recogen los datos (por ejemplo, tipo Likert).

Un modelo que se ajusta a las características de la variable autocontrol es el modelo de escala de clasificación (MEC). Este es un modelo politómico derivado del modelo de Rasch que posee tres importantes supuesto (Andrich, 1978): la unidimensionalidad (todos los ítems deben medir el mismo constructo), la independencia local de los ítems y de las personas (la respuesta de un individuo a cualquiera de los ítems del

test no se verá afectada por su respuesta en otro ítem; y la discriminación homogénea de los elementos (todos los elementos tienen el mismo poder de discriminación).

Cabe mencionar, que la aplicación del MEC requiere que las escalas posean categorías de respuesta gradual, que sean determinadas, exhaustivas y excluyentes. Estas categorías pueden representar cantidades o incrementos en las variables medidas a través de los ítems y las respuestas de las personas indican su grado de acuerdo o desacuerdo con los mismos (Andrich, 1978). Asimismo, el MEC permite estimar la probabilidad de una persona de responder a una categoría determinada en un ítem. Esto resulta de considerar la diferencia entre el nivel de capacidad de la persona en la variable que se está midiendo y el nivel de intensidad o la capacidad de los ítems para medir la variable. Para esto se utiliza el concepto de *item steps*, que se define como el punto en el continuo donde se produce la transición entre dos categorías de respuesta adyacentes (Andrich, 1978).

Por lo tanto, el propósito de la presente investigación fue examinar las propiedades psicométricas de la SCS en una muestra de participantes argentinos para mejorar su comprensión, como así también proporcionar fundamentos para la revisión y mejoramiento de las cualidades de medición de la escala. Para cumplir con este objetivo, se utilizó el MEC y se evaluó la efectividad de las categorías de calificación, la dimensionalidad de la escala y las distribuciones de los índices de dificultad de los ítems y las medidas de habilidad de las personas.

Método

Participantes

Los participantes fueron estudiantes de las facultades de la Universidad Nacional de Córdoba (UNC), Argentina. La muestra estuvo compuesta por 1072 estudiantes de ambos sexos, 791 participantes mujeres (73,8%) y 281 varones (26,2%). Las edades de los participantes se encontraron comprendidas entre los 18 y 25 años ($M = 22,71$; $DS = 3,8$). El tipo de muestreo utilizado fue por conveniencia, ya que el estudio se desarrolló a partir de la participación de aquellos estudiantes que contestaron la encuesta difundida mediante las redes sociales. En cuanto a la distribución de los participantes según la facultad de pertenencia, los porcentajes fueron: 2,1 Artes, 4,0 Ciencias Agropecuarias, 2,0 Arquitectura, Urbanismo y Diseño, 9,2 Derecho y Ciencias Sociales, 4,5 Ciencias Económicas, 10,6 Ciencias Exactas Físicas y Naturales, 2,8 Filosofía y Humanidades, 7 Lenguas, 2,6 Matemática, Astronomía y Física, 10,8 Ciencias Médicas, 1,9 Odontología, 10,5 Psicología, 2,0 Ciencias Químicas. El resto de los participantes indicó que entre las opciones de respuesta no encontraron su facultad.

Instrumento

Self-Control Scale (SCS; Tangney et al., 2004). La SCS permite evaluar la capacidad que tienen los individuos para dejar sin efecto, modificar o dejarse llevar por tendencias internas impulsivas (Tangney et al., 2004). Este instrumento consta de 36 ítems agrupados en cinco factores: Autodisciplina con 9 ítems (e.g., “Soy bueno para resistir tentaciones”), Hábitos Deliberados/No Impulsivos con 10 ítems (e.g., “Digo cosas inapropiadas”), Hábitos Saludables con 7 ítems (e.g., “Me alimento de manera saludable”), Ética de Trabajo con 5 ítems (e.g., “El placer y la diversión a veces me distraen y no hago mi trabajo”) y Confiabilidad con 5 ítems (e.g., “La gente puede contar conmigo en los tiempos acordados”). En la consigna de la escala se solicita a los participantes que indiquen su grado de acuerdo con la afirmación que se presenta en cada ítem. Para ello, utilizan una escala de respuesta tipo Likert con cinco opciones (desde 1 = para nada hasta 5 = *totalmente*). Aunque los ítems se distribuyen en cinco factores, los autores informan la consistencia interna de la SCS en un solo factor, que reportó un alfa de Cronbach de 0,89. Asimismo, los análisis de validez predictiva y discriminante realizados indicaron un buen índice de validez de la escala (Tangney et al., 2004).

Para la traducción de los ítems de la SCS tres expertos de la lengua inglesa realizaron, de manera independiente, una traducción directa de la versión original del inventario al español. Las tres versiones fueron comparadas y sometidas a discusión hasta obtener una versión consensuada de cada uno de los ítems. En este paso se tuvieron en cuenta las diferencias lingüísticas y culturales entre la población de origen y la población destino de esta versión adaptada.

Procedimiento

La administración de la SCS se realizó en formato on-line, mediante la plataforma virtual *Lime Survey*, que ofrece el servidor de encuestas de la UNC. El link del cuestionario se difundió a través de correo electrónico y distintas redes sociales, principalmente Facebook. Se buscó mayor heterogeneidad a partir de la publicación del link en grupos de diversas carreras de distintas facultades de la UNC.

Con respecto a los aspectos formales de la administración, no se habilitó a los participantes la posibilidad de guardar sus respuestas y volver a culminar los cuestionarios en otro momento, como así también se imposibilitó realizar la encuesta más de una vez desde una misma dirección IP, a modo de evitar respuestas duplicadas o más de una respuesta por persona. El tiempo requerido para completar el cuestionario fue de aproximadamente 20 minutos.

Se solicitó a los participantes de la encuesta su consentimiento y se les informó sobre los fines de la investigación y la confidencialidad de los datos obtenidos. Para completar el cuestionario se necesitaba que los participantes confirmaran (mediante un clic en la opción *aceptar*) haber entendido y aceptado participar en el estudio. En caso de que los estudiantes no otorgaran su aprobación, la encuesta no se habilitaba. Los procedimientos del estudio, incluyendo los aspectos éticos, fueron aprobados por la Comisión Evaluadora de Proyectos de la Secretaría de Ciencia y Técnica de la UNC.

Análisis de Datos

El análisis de datos fue realizado utilizando el MEC con el software *Winsteps* (Linacre, 2009). Se eligió el MEC, ya que el formato de respuestas viene definido de la misma forma para todos los ítems de un test (formato tipo *Likert*) y se asume que el valor de cada categoría de respuestas será el mismo y que la única diferencia de los ítems se debe a su diferente localización en el continuo unidimensional de la variable que se mide (Barbero García, 1999). El plan de calibración de los ítems consistió de los siguientes pasos:

Eficacia de las categorías de calificación. Se evaluó la eficacia de la escala de calificación para determinar la categorización óptima. Como paso previo a explorar la escala de calificación, se examinó la correlación biserial puntual para asegurar que todos los elementos estaban orientadas en la misma dirección en la variable latente, es decir, que todas las correlaciones fueran positivas (Linacre, 2002). Luego se examinó la exactitud de la escala de clasificación, siguiendo los siguientes criterios. Primero, que al menos se presenten 10 observaciones por cada categoría de calificación, para poder estimar los valores de los umbrales de forma estable. Segundo, que las medidas promedio (la habilidad promedio de todas las personas que eligieron una categoría en particular) aumenten de forma monótona (en tamaño) a medida que la variable aumenta. Tercero, los umbrales (las dificultades estimadas para elegir una categoría de respuesta sobre otra) también deben aumentar monótonamente a través de la escala de calificación; si no lo hacen, se la considera desordenada. Cuarto, las magnitudes de las distancias entre los umbrales de categoría adyacentes deben indicar que cada paso define un rango distinto en la variable. Linacre (2002) sugiere que los umbrales deben aumentar al menos 1,4 logits para una escala de 3 puntos, 1,1 para una escala de 4 puntos y 0,81 para una escala de 5 puntos para mostrar la distinción entre las categorías, pero no más de 5 logits, a fin de evitar grandes brechas en la variable. Por último, el ajuste estadístico proporciona otro criterio para evaluar la calidad de una escala de calificación. Sin embargo, cuando se obtiene un valor *Outfit* superior a 2 indica más desinformación que información, por lo que la categoría introduce ruido en el proceso de medición. Estos criterios se utilizan generalmente en combinación, para detectar cualquier categorización desordenada, la definición de categorías problemáticas factibles de colapsar y para determinar la categorización óptima.

Unidimensionalidad. La unidimensionalidad de la escala se evaluó mediante el análisis de componentes principales de los residuos y un análisis factorial exploratorio (AFE). Se consideró que se cumple el supuesto de unidimensionalidad si el modelo de medida (el modelo de Rasch unidimensional) explicaba aproximadamente un 50% de la varianza. Si el mayor factor adicional (una dimensión secundaria) tiene un valor propio menor de 3 (una fuerza de tres ítems) y explica menos del 5% de la varianza inexplicada, la unidimensionalidad puede sostenerse (Linacre, 2009). Para el AFE se utilizó el método de Hull para seleccionar el número de factores comunes (Timmerman & Lorenzo-Seva, 2011), mediante el programa FACTOR 9.3 (Lorenzo-Seva & Ferrando, 2013).

Ajuste del modelo de Rasch. Se realizaron tres análisis: el ajuste global de los datos, el ajuste de los ítems y el ajuste de las personas. Con el primero se comprueba si, en términos generales, la matriz de datos se ajusta a lo pronosticado por el modelo. El ajuste de los ítems permite estudiar a cada uno de estos de manera independiente. Con el ajuste de las personas se identifica a los participantes que respondieron de manera incoherente a la formulación teórica. En este caso, los estadísticos de ajuste que se utilizaron para la calibración de ítems fueron los índices Infit (ajuste interno) y Outfit (ajuste externo). Por una parte, el índice Infit es un estadístico de ajuste calculado a partir de las medias cuadráticas sin estandarizar, que permite identificar comportamientos inesperados que afectan a los ítems que, en el continuo de medida, se encuentran cerca del nivel de rasgo que una persona posee. Por otra parte, el índice Outfit se interpreta como la media cuadrática ponderada de residuales resultante de las personas y de los ítems; este índice permite evaluar el comportamiento inesperado de los ítems que tienen una dificultad alejada del nivel de rasgo latente que presenta cada persona (Bond & Fox, 2007). Cuando los datos observados coinciden con los propuestos por el modelo, los valores de Infit y Outfit tienen valores próximos a 1; en caso contrario, se obtendrán valores alejados de 1. Es decir, un valor de Infit de 1 indica que el 100% de la varianza de los datos empíricos es explicado por el modelo, mientras que un valor mayor a 1,3 indica que hay más varianza de la esperada (un 30% de la varianza no puede ser explicada por el modelo). Siguiendo los criterios propuesto por Linacre (2002), la región para considerar un ajuste aceptable oscila entre 0,6 y 1,3 logits.

Separación y confiabilidad. Los ítems deben estar lo suficientemente separados en niveles de dificultad, para poder identificar el sentido y significado de la variable latente (Wright & Stone, 2004). El índice de separación de las personas indica la aptitud del instrumento para discriminar a las personas en la variable medida. Un conjunto útil de ítems debe definir al menos tres estratos de personas (por ejemplo, los niveles altos, moderados y bajos de conocimiento). El índice de separación superior a 2 se considera adecuado (Bond & Fox, 2007), como también una confiabilidad asociada al índice de separación de 0,80 (Bond & Fox, 2003). Además, a través del mapa de ítems y personas se puede ver de manera conjunta cuál es la posición de los ítems y las personas en el continuo.

Funcionamiento diferencial del ítem (DIF). Se realizaron análisis de DIF según el sexo y la edad de los participantes. Un ítem presenta DIF cuando la probabilidad de respuesta correcta no depende únicamente del nivel de la persona en el rasgo intencionadamente medido por el test (Bond & Fox, 2007). Para el análisis de DIF según la edad se clasificaron los participantes en *jóvenes* y *adulto*, mediante el cálculo de la mediana. Para aplicar el DIF se realizaron análisis *pairwise*, fijándose el nivel de significación en $\alpha < 0,01$. Se consideró que el contraste del DIF debía ser $\geq 0,5$ logits para determinar la existencia de diferencias atribuibles, en este caso, a la variable sexo o edad (Linacre, 2006).

Objetividad específica. Se realizó un análisis de la objetividad específica (Rasch, 1977). Una medida solo puede ser considerada válida y generalizable si no depende de las condiciones específicas con que ha sido obtenida. Es decir, la diferencia entre dos personas en un atributo no debe depender de los ítems específicos con los que sea estimada. Igualmente, la diferencia entre dos ítems no debe depender de las personas específicas que se utilicen para cuantificarla. Esta propiedad fue denominada objetividad específica por Rasch (1977). Uno de los principales procedimientos que se recomiendan para analizar el ajuste de los datos al modelo consiste en contrastar empíricamente los supuestos teóricos que el modelo sostiene y los datos obtenidos de la administración del instrumento de evaluación (Hambleton, Swaminathan & Rogers, 1991). Por este motivo, para analizar la objetividad específica se dividió aleatoriamente la muestra de alumnos en dos submuestras con tamaño similar, se estimaron los parámetros de dificultad de los ítems en cada submuestra y se llevó a cabo una regresión lineal simple entre los valores β obtenidos (Prieto & Delgado, 2003). Los valores esperados para la correlación lineal de Pearson entre ambos conjuntos de parámetros, la ordenada en el origen y la pendiente de la recta que indican un ajuste perfecto serían 1, 0 y 1, respectivamente.

Resultados

Eficacia de las Categorías de Calificación

La correlación biserial puntual para los 36 ítems fue positiva, los que indica que cada ítem trabaja en conjunto con la escala a la que pertenece, es decir, los ítems se encuentran en el mismo sentido y midiendo el mismo constructo. Con respecto a la estructura de clasificación, se cumplieron casi todos los criterios

propuestos por Linacre (2002). A nivel general, se observó que: hubo altas frecuencias para todas las categorías; el incremento medio de la medida fue monótono a lo largo de las categorías; los valores de Outfit tienen valores próximos a 1 para todas las categorías y los umbrales también aumentan monótonamente, lo cual estaría indicando que cada categoría es la más probable para un rango específico en el continuo del constructo. Sin embargo, las distancias entre los umbrales consecutivos no son lo suficientemente grandes como para describir gamas distintas en la variable. Por lo tanto, las categorías originales de la escala se colapsaron en categorías adyacentes de diferentes maneras y luego se volvieron a analizar los datos para comparar diferentes categorizaciones de la escala de calificación. La combinación de las categorías 2 y 3 y 4 y 5 (es decir, 12244) fue mejor que otras combinaciones (ver Tabla 1), porque se les ordenó todas las medidas y calibraciones de paso (0,03, 0,61 y 1,39), avanzando lo suficiente (1,64).

Tabla 1
Comparación de las Categorías de Clasificación

Categoría	Recuento observado	Medida promedio	Outfit mean square	Calibración
Original 12345				
1	4946	-0,22	1,05	Ninguno
2	6245	-0,03	1,04	-0,38
3	9211	0,12	0,89	-0,34
4	8706	0,35	1,01	0,31
5	9484	0,66	0,97	0,41
Tres categorías				
12244				
1	4946	0,03	1,02	Ninguno
2	15456	0,61	0,94	-0,82
3	18154	1,39	1,01	0,82

Unidimensionalidad

Primero se examinó la unidimensionalidad del instrumento con la estructura de la escala de repuesta original. El análisis de componentes principales de los residuos mostró que la dimensión de Rasch explicó 37,2% de la varianza en los datos con su valor propio de 21,3. El primer contraste (la mayor dimensión secundaria) tuvo un valor propio de 3,1 y representó el 5,5% de la varianza no explicada. La variación en los datos explicados por las medidas Rasch fue más de cuatro veces la varianza explicada por la mayor dimensión secundaria. Los resultados globales indican que se cumple parcialmente el supuesto de unidimensionalidad. En efecto, no puede excluirse la posibilidad de que también pueda existir una dimensión secundaria, debido a que la mayor dimensión secundaria tenía la fuerza de casi tres artículos y representó un poco más de 5% de la varianza explicada. En el AFE, la medida de adecuación muestral de Kaiser-Meyer-Olkin (KMO) obtenida = 0,847 y la prueba de esfericidad de Bartlett ($gl = 630$, $p = 0,001$) = 7943,9, advirtieron la viabilidad de realizar el análisis factorial. El método de Hull sugirió la extracción de un solo factor (índice de bondad de ajuste = 0,90).

El análisis de componentes principales del residual de la estructura de tres categorías arrojó resultados similares a la de la estructura de cinco categorías originales. En concreto, el 35,8 % de la varianza (valor propio de 20,1) se explica por la dimensión de Rasch en los datos y la dimensión secundaria tenía un valor propio de 2,8 y representó el 5% de la varianza no explicada. El KMO = 0,843 y la prueba de esfericidad de Bartlett ($gl = 630$, $p = 0,001$) = 6531,5, advirtieron la viabilidad de realizar el análisis factorial; el método de Hull sugirió la extracción de un solo factor (Índice de Bondad de Ajuste = 0,91).

Ajuste del Modelo de Rasch

El ajuste de los ítems fue adecuado para 35 ítems cuando se utilizó la escala original. Solo el ítem I20 (No impulsivos) presentó un ajuste inadecuado, Infit MNSQ = 1,51, Outfit MNSQ = 1,57. La medida de

($DE = 0,71$). Los valores de Infit de los ítems variaron entre 0,77 y 1,39 y Outfit, entre 0,77 y 1,44, con una media de 1,01 ($DE = 0,16$) y 1,02 ($DE = 0,17$), respectivamente. El análisis de ajuste de las personas refleja que el 87,80% de los patrones de respuesta se ajustó al modelo (Infit $\geq 0,60$ y Outfit $\leq 1,40$). Los niveles de habilidad variaron entre $-1,10 \leq \theta \leq 3,45$, con una media de 0,90 ($DE = 0,70$) (ver Tabla 2). La distribución de los ítems y personas en el mapa de Wright para la escala de tres opciones es semejante a la de la escala original.

Separación y Confiabilidad

El índice de separación de los ítems (14,99), el índice de confiabilidad de los ítems (1,00), los valores de los índices de separación de persona (2,40) y confiabilidad de las personas (0,85) fueron satisfactorios con la escala original. Estos resultados indican que la muestra utilizada es suficientemente grande como para confirmar la jerarquía de dificultad del ítem (validez de constructo) del instrumento (Linacre, 2006). Por otro lado, el índice de separación de los ítems (13,42), el índice de confiabilidad de los ítems (0,99) y los valores de los índices de separación de persona (2,22) y confiabilidad de las personas (0,83) fueron satisfactorios y levemente inferiores a la escala original cuando se utilizó la escala de tres opciones.

Funcionamiento Diferencial del Ítem

En el análisis de DIF en la escala original según el sexo y la edad de los participantes, se observó que 9 ítems según la edad y 1 según el sexo fueron estadísticamente significativos ($p = 0,001$), pero el valor de contraste no era mayor a 0,50. En el caso de la escala de tres opciones, en 13 ítems se observaron diferencias significativas, pero solo se encontraron diferencias de acuerdo al sexo en los ítems I3 (No impulsivos), I5 (Hábitos Saludables) e I36 (Confiabilidad), en donde los contrastes fueron superiores a 0,50. En el ítem I3 la dificultad del ítem (media de DIF) para la muestra femenina fue de -0,79 logits, mientras que para la masculina fue de -0,27, en el ítem I5 fue de -1,15 logits para el grupo de mujeres y de -0,63 para el de hombres y en el ítem I36 fue de -1,98 para las mujeres y -2,86 para los hombres. A partir de estos valores, es posible indicar que los ítems I3 e I5 fueron más difíciles para las mujeres, mientras que en el ítem I36 la dificultad mayor la presentaron los hombres.

Objetividad Específica

Los resultados mostraron un valor de $r = 0,99$, el valor de la constante fue 0,001 y $\beta = 0,99$ para la escala original y de $r = 0,98$, el valor de la constante 0,01 y $\beta = 0,98$ para la escala con tres opciones. Estos resultados permiten asumir que se cumple satisfactoriamente el requisito de objetividad específica y que los datos presentan un buen ajuste global al modelo (Prieto & Delgado, 1999).

Discusión

En este estudio se presentan los resultados obtenidos a partir de la aplicación del modelo de Rasch, a saber, el MEC, el cual se utilizó para valorar las propiedades psicométricas de la Escala de Autocontrol (SCS). El MEC, como una extensión del modelo de Rasch para ítems politómicos, permite realizar una medición conjunta, con estadísticos suficientes, objetividad específica, propiedades de intervalo, especificidad del error típico de medida y el ajuste de los patrones de respuesta de los participantes (Prieto & Delgado, 2003). En líneas generales, puede decirse que la escala SCS presentó adecuadas propiedades psicométricas.

Respecto del ajuste de los datos al modelo, se observó que en su gran mayoría los participantes respondieron coherentemente a los ítems del test, permitiendo así, identificar patrones de respuestas predecibles por el modelo propuesto por la SCS (Linacre, 2002). Por lo que refiere al ajuste global de los ítems, este fue adecuado para 35 de los 36 ítems que componen la escala. Si bien existe un único ítem que no se ajusta a lo que el modelo podría predecir, cabe mencionar que en la literatura no se encuentran motivos teóricos que expliquen tal comportamiento (Hasford & Bradley, 2011).

Por su parte, el mapa de distribución de ítems-personas, utilizando tanto las categorías originales como las propuestas en este trabajo, permitió comparar los índices de dificultad y los niveles de habilidad de los respondientes. Desde este enfoque se puede analizar la validez del constructo, ya que los ítems que componen el instrumento deben estar distribuidos de forma creciente en dificultad, como también deben representar

adecuadamente los contenidos de interés. La ausencia de ítems para cada nivel de habilidad marca la necesidad de ampliar el número de ítems para lograr una mejor cobertura de la prueba.

Tabla 2
Ajuste de los Ítems, Medida y Personas

Self-Control Scale	Medida	Desviación estándar	MNSQ		r_{bp}
			Ajuste próximo	Ajuste lejano	
<i>Resumen de las personas</i>					
Media	0,24	0,15	1,01	1,03	
Desviación estándar	0,39	0,01	0,35	0,40	
Máximo	2,57	0,43	3,30	4,21	
Mínimo	-1,00	0,14	0,30	0,31	
<i>Resumen de los ítems</i>					
Media	0,00	0,03	1,02	1,03	
Desviación estándar	0,41	0,00	0,20	0,21	
Máximo	0,78	0,04	1,51	1,57	
Mínimo	-1,25	0,03	0,74	0,75	
I1 Autodisciplina	0,11	0,03	0,85	0,85	0,40
I2 No impulsivos	0,33	0,03	0,74	0,75	0,47
I3 No impulsivos	-0,28	0,03	0,80	0,80	0,39
I4 No impulsivos	0,12	0,03	1,10	1,15	0,18
I5 Hábitos Saludables	-0,54	0,03	1,02	1,01	0,39
I6 Confiabilidad	-0,81	0,03	1,27	1,39	0,19
I7 Ética de Trabajo	0,59	0,03	1,21	1,22	0,36
I8 Autodisciplina	0,15	0,03	1,22	1,22	0,26
I9 Autodisciplina	-0,10	0,03	0,89	0,88	0,42
I10 No impulsivos	0,04	0,03	1,01	1,03	0,26
I11 No impulsivos	-0,29	0,03	1,14	1,11	0,43
I12 Hábitos Saludables	-0,31	0,03	1,11	1,12	0,30
I13 Hábitos Saludables	-0,11	0,03	1,02	1,02	0,35
I14 Confiabilidad	0,18	0,03	1,00	1,03	0,24
I15 Ética de Trabajo	0,50	0,03	0,80	0,84	0,18
I16 Autodisciplina	0,60	0,03	1,28	1,27	0,36
I17 Autodisciplina	0,16	0,03	0,83	0,84	0,36
I18 Autodisciplina	0,52	0,03	0,89	0,92	0,31
I19 No impulsivos	0,07	0,03	0,82	0,82	0,52
I20 No impulsivos	-0,64	0,03	1,51	1,57	0,18
I21 Hábitos Saludables	0,38	0,03	0,86	0,88	0,25
I22 Ética de Trabajo	0,78	0,03	1,26	1,26	0,33
I23 Autodisciplina	0,11	0,03	1,10	1,15	0,22
I24 No impulsivos	0,19	0,03	1,00	1,01	0,41
I25 Hábitos Saludables	-0,02	0,03	0,95	0,98	0,23
I26 Hábitos Saludables	0,01	0,03	0,83	0,85	0,28
I27 Ética de Trabajo	0,32	0,03	0,75	0,75	0,51
I28 Autodisciplina	0,29	0,03	0,83	0,83	0,48
I29 Confiabilidad	-0,32	0,03	0,97	1,00	0,31
I30 Autodisciplina	0,04	0,03	0,89	0,88	0,46
I31 No impulsivos	-0,12	0,03	0,78	0,77	0,50
I32 No impulsivos	0,08	0,03	1,11	1,11	0,33
I33 No impulsivos	-0,22	0,03	0,93	0,93	0,34
I34 Hábitos Saludables	-0,56	0,03	1,39	1,39	0,29
I35 Confiabilidad	0,01	0,03	1,26	1,26	0,26
I36 Confiabilidad	-1,25	0,04	1,30	1,35	0,14

Nota. r_{bp} : correlación biserial puntual.

Cabe mencionar que las distribuciones de ambos parámetros (habilidad de las personas y dificultad de los ítems) cubrieron gran parte del continuo autocontrol, aunque la mayoría de los reactivos de la escala se localizaron en la zona media del mapa de distribución de ítems-personas. Por ende, no se observó la presencia de reactivos en el extremo superior e inferior del continuo. Esta distribución de los ítems indicaría que la SCS no permitiría valorar niveles muy altos o muy bajos de autocontrol, lo cual coincide con otros estudios que han reportado resultados similares (Hasford & Bradley, 2011). Considerar esto resulta de interés, ya que diversos estudios que utilizan la SCS llegan a conclusiones que implican afirmaciones respecto a altos niveles o bajos niveles de autocontrol. Como puede apreciarse, estas inferencias no serían adecuadas, puesto que la escala no permitiría la valoración de intensidades extremas del atributo (Hasford & Bradley, 2011).

Respecto de la precisión de la escala, se observó que el error de estimación de todos los ítems es relativamente bajo. De igual forma, si se presta atención al índice de confiabilidad de las personas y de los ítems, se aprecia que ambos son muy elevados, lo que indica que la localización de las personas e ítems sería previsiblemente reproducible (Andrich, 1988). Tomando estos resultados y siguiendo lo propuesto por Messick (1989) y Bond (2004) respecto al concepto de validez, podría afirmarse que al encontrar un ajuste adecuado, por ejemplo, de los índices de separación de los ítems, confiabilidad de los ítems y los índices de separación de persona y confiabilidad de las personas, sería posible inferir que los ítems se comportan según las predicciones del modelo psicométrico, lo cual aporta evidencia a favor de la validez de constructo (Bond, 2004).

En este estudio el modelo de medida explicó el 37,2% de la varianza y el AFE, utilizando el método de Hull, sugirió la extracción de un solo factor. En principio, considerando los datos de la varianza explicada, podría decirse que uno de los supuestos más importantes del modelo de Rasch no cumple con los criterios estadísticos requeridos para considerar la plena unidimensionalidad del instrumento. Sin embargo, cabe mencionar que la posibilidad de que un solo factor explique por completo la varianza total de las puntuaciones resulta de gran complejidad (Muñiz Fernández, 1997). Asimismo, Muñiz Fernández (1997) afirmó que por este motivo “la unidimensionalidad se convierte en una cuestión de grado, cuanta más varianza explique el primer factor más unidimensionalidad existirá” (p. 26), es decir, considerando que la unidimensionalidad es relativa al grado, el porcentaje no necesariamente indica ausencia de unidimensionalidad (Muñiz Fernández, 1997). Por este motivo, considerando lo propuesto por este autor y los resultados del AFE, se consideró que el constructo autocontrol era lo suficientemente unidimensional para realizar el análisis de Rasch.

Respecto de la estructura original de calificación de cinco categorías de respuesta, se observó que no funcionó con total eficacia, ya que las distancias entre los umbrales consecutivos no fueron lo suficientemente amplios como para describir gamas distintas en la variable medida. De esta manera, a partir de los análisis ad hoc se encontró que la medición con tres categorías de respuesta podría ser más efectiva que con cinco. Este último resultado en relación al cambio en las categorías de respuesta conlleva que se realice una nueva administración del instrumento. Cabe mencionar que, debido a la dificultad para obtener una nueva muestra, no se realizó un estudio que permitiera poner a prueba la efectividad de esta recomendación. Sin embargo, en el presente estudio se decidió presentar una posible modificación al instrumento, con el fin de brindar una alternativa para mejorar las propiedades de medición de la escala.

En lo que concierne al funcionamiento de los ítems de la escala, se realizaron análisis de DIF según el sexo y la edad de los participantes, observándose que en la escala original solo 10 ítems presentaron diferencias estadísticamente significativas; sin embargo, el valor de contraste no era mayor al punto de corte recomendado ($\geq 0,5$ logits) (Linacre, 2006). Por su parte, en la escala de tres opciones, en 13 ítems se observaron diferencias significativas, pero solo en tres los contrastes fueron superiores al punto de corte. Como las diferencias solo se encontraron en la versión de la escala con tres categorías de respuesta, las implicancias que pueden tener estos resultados exceden los objetivos del presente trabajo. No obstante, sería recomendable poner a consideración de expertos en la temática aquellos ítems que presentaron diferencias estadísticamente significativas para determinar las causas que influyen en el sesgo de estos ítems.

En base a los resultados hasta aquí expuestos, en futuras investigaciones se deberían realizar estudios de validez predictiva que permitan valorar la capacidad del instrumento para anticipar comportamientos vinculados al autocontrol, como el bajo rendimiento académico, el comportamiento violento, el embarazo adolescente, entre otros comportamientos vinculados a la variable (Meece & Robinson, 2014; Perrone, Sullivan, Pratt & Margaryan, 2004; Reisig, Wolfe & Pratt, 2012; Wolfe, 2015). Asimismo, se debería realizar un nuevo estudio con las tres categorías propuestas en el presente estudio y agregar nuevos ítems que permitan medir valores extremos de la variable autocontrol, es decir, tanto niveles muy bajos como muy elevados de autocontrol, debido a que en la literatura se reconoce que existe una asociación diferenciada para

cada uno de los niveles de autocontrol y determinados problemas de comportamiento (Tangney et al., 2004). Asimismo, la redacción de nuevos reactivos favorecería que el porcentaje de varianza explicada por una dimensión aumentara, dado que se estima que los reducidos porcentajes de varianza explicada en el presente artículo por la dimensión autocontrol se debe a la reducida dispersión en la dificultad de los ítems, ya que la varianza explicada depende conjuntamente de la dispersión de personas e ítems.

A modo de conclusión, el presente estudio demuestra que la utilización del modelo de Rasch resulta útil para la revisión y mejora de la SCS, mediante la utilización de una muestra por conveniencia constituida por estudiantes universitarios. Además, esta investigación ofrece información detallada sobre la metodología idónea para los estudios futuros que pretendan replicar los resultados del presente trabajo, como así también seguir avanzando en el estudio de las propiedades psicométricas del instrumento (mediante la TRI). También se provee una escala de respuesta alternativa a la original, que permitiría obtener con mayor claridad información sobre el constructo autocontrol. Para finalizar, cabe mencionar que la utilización de la escala en estudiantes universitarios debería ser utilizada con fines de investigación y, con fines prácticos, solo cuando se pretenda valorar niveles moderados de autocontrol.

Referencias

- Andrich, D. (1978). Application of a psychometric rating model to ordered categories which are scored with successive integers. *Applied Psychological Measurement*, 2, 581-594. <https://doi.org/10.1177/014662167800200413>
- Andrich, D. (1988). *Rasch models for measurement*. Newbury Park, CA: SAGE.
- Barbero García, M. I. (1999). Desarrollos recientes de los modelos psicométricos de la teoría de respuesta a los ítems. *Psicothema*, 11, 195-210.
- Bond, T. G. (2004). Validity and assessment: A Rasch measurement perspective. *Metodología de las Ciencias del Comportamiento*, 5, 181-196.
- Bond, T. G. & Fox, C. M. (2003). Applying the Rasch model: Fundamental measurement in the human sciences. *Journal of Educational Measurement*, 40, 185-187. <https://doi.org/10.1111/j.1745-3984.2003.tb01103.x>
- Bond, T. G. & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2ª ed.). Mahwah, NJ: Lawrence Erlbaum.
- de Ridder, D. T. D., Lensvelt-Mulders, G., Finkenauer, C., Stok, F. M. & Baumeister, R. F. (2012). Taking stock of self-control: A meta-analysis of how trait self-control relates to a wide range of behaviors. *Personality and Social Psychology Review*, 16, 76-99. <https://doi.org/10.1177/1088868311418749>
- Dixon, M. R. & Tibbetts, P. A. (2009). The effects of choice on self-control. *Journal of Applied Behavior Analysis*, 42, 243-252. <https://doi.org/10.1901/jaba.2009.42-243>
- Duckworth, A. L. & Kern, M. L. (2011). A meta-analysis of the convergent validity of self-control measures. *Journal of Research in Personality*, 45, 259-268. <https://doi.org/10.1016/j.jrp.2011.02.004>
- Embretson, S. E. & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum.
- Grasmick, H. G., Tittle, C. R., Bursik, R. J. & Arneklev, B. J. (1993). Testing the core empirical implications of Gottfredson and Hirschi's general theory of crime. *Journal of Research in Crime and Delinquency*, 30, 5-29. <https://doi.org/10.1177/0022427893030001002>
- Hambleton, R. K. (2000). Response to Hays et al and McHorney and Cohen: Emergence of item response modeling in instrument development and data analysis. *Medical Care*, 38(Supplement II 9), II60-II65.
- Hambleton, R. K. & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston, MA: Kluwer-Nijhoff.
- Hambleton, R. K., Swaminathan, H. & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: SAGE.
- Hasford, J. & Bradley, K. D. (2011). Validating measures of self control via Rasch measurement. *Journal of Applied Business Research*, 27(6), 45-56. <https://doi.org/10.19030/jabr.v27i6.6465>
- Jones, L. V. & Thissen, D. (2007). A history and overview of psychometrics. En C. R. Rao & S. Sinharay (Eds.), *Handbook of statistics 26: Psychometrics* (pp. 1-27). Amsterdam, Holanda: Elsevier.
- Linacre, J. M. (2002). Understanding Rasch measurement: Optimizing rating scale category effectiveness. *Journal of Applied Measurement*, 3, 85-106.
- Linacre, J. M. (2006). *Winsteps® Rasch measurement computer program*. Beaverton, OR: Winsteps.com
- Linacre, J. M. (2009). *Winsteps® (Version 3.68. 0) [Software computacional]*. Beaverton, OR: Winsteps.com
- Lindner, C., Nagy, G. & Retelsdorf, J. (2015). The dimensionality of the Brief Self-Control Scale—an evaluation of unidimensional and multidimensional applications. *Personality and Individual Differences*, 86, 465-473. <https://doi.org/10.1016/j.paid.2015.07.006>
- Lorenzo-Seva, U. & Ferrando, P. J. (2013). FACTOR 9.2: A comprehensive program for fitting exploratory and semiconfirmatory factor analysis and IRT models. *Applied Psychological Measurement*, 37, 497-498. <https://doi.org/10.1177/0146621613487794>
- Ludwig, V. U., Stelzel, C., Krutiak, H., Prunke, C. E., Steimke, R., Paschke, L. M. ... Walter, H. (2013). Impulsivity, self-control, and hypnotic suggestibility. *Consciousness and Cognition*, 22, 637-653. <https://doi.org/10.1016/j.concog.2013.04.001>
- Maloney, P. W., Grawitch, M. J. & Barber, L. K. (2012). The multi-factor structure of the Brief Self-Control Scale: Discriminant validity of restraint and impulsivity. *Journal of Research in Personality*, 46, 111-115. <https://doi.org/10.1016/j.jrp.2011.10.001>
- Malouf, E., Stuewig, J. & Tangney, J. (2012). Self-control and jail inmates' substance misuse post-release: Mediation by friends' substance use and moderation by age. *Addictive Behaviors*, 37, 1198-1204. <https://doi.org/10.1016/j.addbeh.2012.05.013>
- Meece, D. & Robinson, C. M. (2014). Father-child interaction: Associations with self-control and aggression among 4.5-year-olds. *Early Child Development and Care*, 184, 783-794. <https://doi.org/10.1080/03004430.2013.818990>
- Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher*, 18(2), 5-11. <https://doi.org/10.3102/0013189X018002005>

- Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, 23(2), 13-23. <https://doi.org/10.3102/0013189X023002013>
- Moon, B., McCluskey, J. D., Blurton, D. & Hwang, H. -W. (2014). Parent and teacher practices as sources of low self-control: Evidence from Korea. *Youth Violence and Juvenile Justice*, 12, 167-187. <https://doi.org/10.1177/1541204013481982>
- Morean, M. E., DeMartini, K. S., Leeman, R. F., Pearlson, G. D., Anticevic, A., Krishnan-Sarin, S. ... O'Malley, S. S. (2014). Psychometrically improved, abbreviated versions of three classic measures of impulsivity and self-control. *Psychological Assessment*, 26, 1003-1020. <https://doi.org/10.1037/pas0000003>
- Muñiz Fernández, J. (1997). *Introducción a la teoría de respuesta a los ítems*. Madrid, España: Pirámide.
- Patton, J. H., Stanford, M. S. & Barratt, E. S. (1995). Factor structure of the Barratt Impulsiveness Scale. *Journal of Clinical Psychology*, 51, 768-774. [https://doi.org/10.1002/1097-4679\(199511\)51:6<768::AID-JCLP2270510607>3.0.CO;2-1](https://doi.org/10.1002/1097-4679(199511)51:6<768::AID-JCLP2270510607>3.0.CO;2-1)
- Perrone, D., Sullivan, C. J., Pratt, T. C. & Margaryan, S. (2004). Parental efficacy, self-control, and delinquency: A test of a general theory of crime on a nationally representative sample of youth. *International Journal of Offender Therapy and Comparative Criminology*, 48, 298-312. <https://doi.org/10.1177/0306624X03262513>
- Prado, C. E., Treeby, M. S. & Crowe, S. F. (2015). Examining relationships between facial emotion recognition, self-control, and psychopathic traits in a non-clinical sample. *Personality and Individual Differences*, 80, 22-27. <https://doi.org/10.1016/j.paid.2015.02.013>
- Prieto, G. & Delgado, A. R. (1999). Medición cognitiva de las aptitudes. En J. Olea Díaz, V. Ponsoda Gil & G. Prieto (Coords.), *Tests informatizados: fundamentos y aplicaciones* (pp. 207-226). Madrid, España: Pirámide.
- Prieto, G. & Delgado, A. R. (2003). Análisis de un test mediante el modelo de Rasch. *Psicothema*, 15, 94-100.
- Rasch, G. (1977). On specific objectivity: An attempt at formalizing the request for generality and validity of scientific statements. *Danish Yearbook of Philosophy*, 14, 58-93.
- Reisig, M. D., Wolfe, S. E. & Pratt, T. C. (2012). Low self-control and the religiosity-crime relationship. *Criminal Justice and Behavior*, 39, 1172-1191. <https://doi.org/10.1177/0093854812442916>
- Tangney, J. P., Baumeister, R. F. & Boone, A. L. (2004). High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. *Journal of Personality*, 72, 271-324. <https://doi.org/10.1111/j.0022-3506.2004.00263.x>
- Timmerman, M. E. & Lorenzo-Seva, U. (2011). Dimensionality assessment of ordered polytomous items with parallel analysis. *Psychological Methods*, 16, 209-220. <https://doi.org/10.1037/a0023353>
- Vohs, K. D., Finkenauer, C. & Baumeister, R. F. (2011). The sum of friends' and lovers' self-control scores predicts relationship quality. *Social Psychological and Personality Science*, 2, 138-145. <https://doi.org/10.1177/1948550610385710>
- Wolfe, S. E. (2015). Low self-control, gender, race, and offending in late life. *Psychology, Crime & Law*, 21, 426-451. <https://doi.org/10.1080/1068316X.2014.989169>
- Wright, B. D. & Stone, M. H. (2004). *Making measure*. Chicago, IL: Phaneron Press.

Fecha de recepción: Noviembre de 2015.

Fecha de aceptación: Mayo de 2017.